

Computational Statistical Physics

Part I: Statistical Physics and Phase Transitions

Oded Zilberberg

February 21, 2020

Contents

1.1	Introduction	1-3
1.2	Classical Statistical Mechanics	1-4
1.2.1	Phase Space	1-5
1.2.2	Liouville Theorem	1-5
1.2.3	Thermal Equilibrium	1-6
1.2.4	Ensembles	1-6
1.2.5	Microcanonical Ensemble	1-7
1.2.6	Canonical Ensemble	1-7
1.3	Ising Model	1-8
1.3.1	Order Parameter	1-10
1.3.2	Fluctuations	1-11
1.3.3	Correlation Length	1-12
1.3.4	Critical Exponents and Universality	1-13
1.4	Monte Carlo Methods	1-14
1.4.1	Markov Chains	1-14
1.4.2	$M(RT)^2$ Algorithm	1-16
1.4.3	Glauber Dynamics (Heat Bath Dynamics)	1-17
1.4.4	Binary Mixtures and Kawasaki Dynamics	1-19
1.4.5	Creutz Algorithm	1-19
1.4.6	Q2R	1-20
1.4.7	Boundary Conditions	1-21
1.4.8	Temporal Correlations	1-21
1.4.9	Decorrelated Configurations	1-23
1.5	Finite Size Methods	1-24
1.5.1	Binder Cumulant	1-25
1.5.2	First Order Transition	1-27

1.6	<i>Cluster Algorithms</i>	1-28
1.6.1	<i>Potts Model</i>	1-28
1.6.2	<i>The Kasteleyn and Fortuin Theorem</i>	1-29
1.6.3	<i>Coniglio-Klein Clusters</i>	1-31
1.6.4	<i>Swendsen-Wang Algorithm</i>	1-31
1.6.5	<i>Wolff Algorithm</i>	1-32
1.6.6	<i>Other Ising-like Models</i>	1-32
1.7	<i>Histogram Methods</i>	1-34
1.7.1	<i>Broad Histogram Method</i>	1-35
1.7.2	<i>Flat Histogram Method</i>	1-36
1.7.3	<i>Umbrella Sampling</i>	1-36
1.8	<i>Renormalization Group</i>	1-37
1.8.1	<i>Real Space Renormalization</i>	1-37
1.8.2	<i>Renormalization and Free Energy</i>	1-38
1.8.3	<i>Majority Rule</i>	1-38
1.8.4	<i>Decimation of the One-dimensional Ising Model</i>	1-39
1.8.5	<i>Generalization</i>	1-41
1.8.6	<i>Monte Carlo Renormalization Group</i>	1-42
1.9	<i>Boltzmann Machine</i>	1-42
1.9.1	<i>Hopfield Network</i>	1-43
1.9.2	<i>Boltzmann Machine Learning</i>	1-44
1.10	<i>Parallelization</i>	1-46
1.10.1	<i>Multi-Spin Coding</i>	1-46
1.10.2	<i>Vectorization</i>	1-47
1.10.3	<i>Domain Decomposition</i>	1-48
1.11	<i>Non-Equilibrium Systems</i>	1-49
1.11.1	<i>Directed Percolation</i>	1-49
1.11.2	<i>Kinetic Monte Carlo (Gillespie Algorithm)</i>	1-50

1.1 Introduction

The lecture is divided into two parts. In the first part, we are going to cover computational approaches to study phase transitions occurring in different statistical physics models including magnetic systems, binary mixtures or models of opinion formation. However, the range of possible applications is much broader. In his book “Phase Transitions”, Ricard V. Solé describes how the study of phase transitions also helps to understand many phenomena observable in biological and ecological complex systems [1]. However, only for a limited number of systems it is possible to derive analytical solutions. For that reason, computational methods have become invaluable to obtain further insights where analytical approaches fail. As an application that became more relevant recently, we also describe how statistical physics forms the basis of many important concepts in the area of machine learning.

The second part focuses on simulation methods of molecular dynamics and establishes a connection between the study of the microscopic interactions of particles and their emerging macroscopic properties that can be analyzed statistically. In particular, we discuss different numerical solution approaches and also introduce methods to simulate molecular interactions in an environment of constant temperature. Moreover, we introduce event-driven algorithms as a more efficient way to simulate molecular systems and finally incorporate quantum mechanical effects in the context of the Car-Parinello method.

Most of the side notes are intended to provide additional information for the interested reader, and to give some relevant examples of current research directions. I want to thank Lucas Böttcher for providing these lecture notes, which are based on previous efforts to summarize the Computational Statistical Physics lecture of Hans J. Herrmann. Comments and questions are always welcome and should be sent to odedz@ethz.ch.

1.2 Classical Statistical Mechanics



Figure 1.1: The importance of statistical physics: At Dufourstrasse 23 in Zürich one can find Ludwig Boltzmann's H-function. This art installation was created by Liam Gillick © Ringier AG.

The field of statistical physics provides methods to study the macroscopic properties of a system consisting of interacting microscopic units. One important example of the achievements of statistical mechanics, are the microscopic theories of thermodynamics of Boltzmann and Gibbs [2, 3].

Our goal is to develop methods which allow us to computationally study phase transitions emerging in different models. To do so, we first have to introduce important concepts from classical statistical mechanics. In the subsequent sections, we only provide a brief summary of these concepts. For those interested in a more detailed discussion of statistical mechanics, we recommend the lectures *Theorie der Wärme* and *Statistical Physics*¹.

In this section, we focus on the treatment of equilibrium systems which exhibit no time dependence. Keeping certain external parameters constant, the notion of a statistical ensemble enables us to interpret macroscopic physical quantities as averages over a large number of such systems in different micro states. However, phase transitions do not only occur in equilibrium systems but also in non-equilibrium models. In recent years, the study of non-equilibrium dynamics became more and more relevant both theoretically and experimentally. We therefore briefly discuss methods have to be employed for studying non-equilibrium phase transitions [4].

¹ The course *Theorie der Wärme* is offered every spring semester in the BSc Physics curriculum and is only taught in German. *Statistical Physics* is a MSc Physics course offered in the autumn term and taught in English.

1.2.1 Phase Space

Let us consider a classical physical system with N particles whose canonical coordinates and the corresponding conjugate momenta are given by q_1, \dots, q_{3N} and p_1, \dots, p_{3N} , respectively. The $6N$ -dimensional space Γ defined by the last set of coordinates defines the *phase space*. This concept has been introduced by Ludwig Boltzmann whose portrait is shown in Fig. 1.2. The considered N particles could simply be uncoupled harmonic oscillators. Then, the phase space of each single particle would look like the one shown in Fig. 1.3. Keeping certain external parameters such as temperature or pressure constant, we could measure a macroscopic physical quantity by computing the time average over different observations of the underlying microscopic states. Since this would involve a cumbersome treatment of the time evolution of all microscopic states, another possibility is to replace the time average by an average over an ensemble of systems in different micro states under the same macroscopic conditions.

The assumption that all states in an ensemble are reached by the time evolution of the corresponding system is referred to as *ergodicity hypothesis*. We define the *ensemble average* of a quantity $Q(p, q)$ as

$$\langle Q \rangle = \frac{\int Q(p, q) \rho(p, q) dpdq}{\int \rho(p, q) dpdq}, \quad (1.1)$$

where $\rho(p, q)$ denotes the *phase space density* and $dpdq$ is a shorthand notation for $dp^{3N}dq^{3N}$.

1.2.2 Liouville Theorem

The dynamics of the considered N particles is described by their *Hamiltonian* $\mathcal{H}(p, q)$, i.e., the equations of motions are

$$\dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} \quad \text{and} \quad \dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i} \quad (i = 1, \dots, 3N). \quad (1.2)$$

Moreover, the temporal evolution of a phase space element of volume V and boundary ∂V is given by

$$\frac{\partial}{\partial t} \int_V \rho dV + \int_{\partial V} \rho v dA = 0, \quad (1.3)$$

where $v = (\dot{p}_1, \dots, \dot{p}_{3N}, \dot{q}_1, \dots, \dot{q}_{3N})$ is a generalized velocity vector.

Applying the divergence theorem to Eq. (1.3), we find that ρ satisfies the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0, \quad (1.4)$$

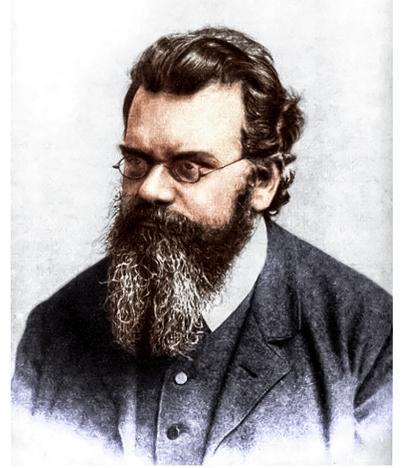


Figure 1.2: Ludwig Boltzmann (1844-1906) is one of the fathers of statistical mechanics and formulated one version of the ergodicity hypothesis.

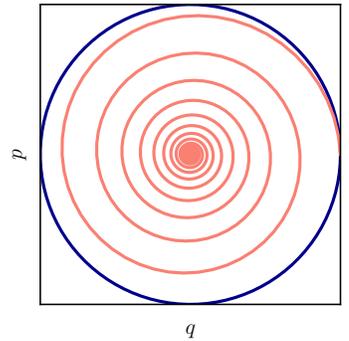


Figure 1.3: Phase spaces of an undamped (blue) and a damped (orange) harmonic oscillator.

where $\nabla = (\partial/\partial p_1, \dots, \partial/\partial p_{3N}, \partial/\partial q_1, \dots, \partial/\partial q_{3N})$. We can further simplify Eq. (1.4) since

$$\nabla \cdot v = \sum_{i=1}^{3N} \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) = \sum_{i=1}^{3N} \underbrace{\left(\frac{\partial}{\partial q_i} \frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial}{\partial p_i} \frac{\partial \mathcal{H}}{\partial q_i} \right)}_{=0} = 0. \quad (1.5)$$

Rewriting Eq. (1.4) using Poisson brackets² yields *Liouville's Theorem*

$$\frac{\partial \rho}{\partial t} = \{\mathcal{H}, \rho\} \quad (1.7)$$

which describes the time evolution of the phase space density ρ .

² The Poisson bracket is defined as

$$\{u, v\} = \sum_i \left(\frac{\partial u}{\partial q_i} \frac{\partial v}{\partial p_i} - \frac{\partial u}{\partial p_i} \frac{\partial v}{\partial q_i} \right) = -\{v, u\}. \quad (1.6)$$

1.2.3 Thermal Equilibrium

In *thermal equilibrium*, the system reaches a steady state in which the distribution of the configurations is constant and time-independent, i.e., $\partial \rho / \partial t = 0$. Liouville's theorem leads to the following condition

$$v \cdot \nabla \rho = \{\mathcal{H}, \rho\} = 0. \quad (1.8)$$

The last equation is satisfied if ρ depends on quantities which are conserved during the time evolution of the system. We then use such a phase space density ρ to replace the *time average*

$$\langle Q \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T Q(p(t), q(t)) dt. \quad (1.9)$$

by its ensemble average as defined by Eq. (1.1).

In the subsequent sections, we also consider discrete configurations X . In this case, we define the ensemble average as

$$\langle Q \rangle = \frac{1}{\Omega} \sum_X Q(X) \rho(X), \quad (1.10)$$

where Ω is the normalizing volume such that $\Omega^{-1} \sum_X \rho(X) = 1$. With the help of ensemble averages, systems can be described by means of some macroscopic quantities, such as temperature, energy and pressure.

1.2.4 Ensembles

We perform measurements to determine some observable quantities characterizing a given system while keeping other parameters constant. This agrees well with the idea behind the definition of ensemble averages, i.e., having a phase space density ρ which only depends on conserved quantities. However, in general, it is not possible to independently adjust the values of all parameters. As an intuitive example,

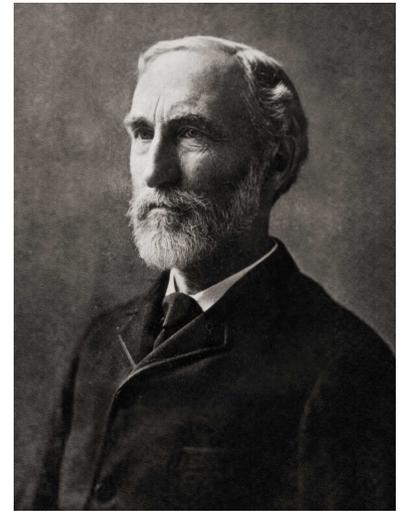


Figure 1.4: Josiah W. Gibbs (1839-1903) introduced the concept of statistical ensembles and derived the laws of thermodynamics from statistical mechanics.

we could consider a classical gas. The dynamics is described by perfect elastic collisions and therefore it is impossible to compress the volume by keeping pressure and temperature unchanged. The system behaves differently depending on which quantities are kept constant. Some quantities such as volume V and pressure p are conjugate to another. Either V can be held constant or p , not both. Other examples are energy E and temperature T , particle number N and chemical potential μ , magnetization M and magnetic field H . Depending on the parameters held constant, the system is described by a

- Microcanonical ensemble: constant E, V, N
- Canonical ensemble: constant T, V, N
- Canonical pressure ensemble: constant T, p, N
- Grandcanonical ensemble: constant T, V, μ

The notion of a physical ensemble has been introduced by Josiah W. Gibbs who is shown in Fig. 1.4. In the subsequent sections, we consider microcanonical and canonical systems and therefore also restrict our discussion of ensembles to these two cases.

1.2.5 Microcanonical Ensemble

The *microcanonical ensemble* is defined by a constant number of particles, volume and energy. Thus, any configuration X of the system has the same energy $E(X) = \text{const}$. The phase space density is also constant and given by

$$\rho(X) = \frac{1}{Z_{\text{mc}}} \delta(\mathcal{H}(X) - E), \quad (1.11)$$

with Z_{mc} being the *partition function* of the microcanonical ensemble

$$Z_{\text{mc}} = \sum_X \delta(\mathcal{H}(X) - E).$$

1.2.6 Canonical Ensemble

Microcanonical ensembles are difficult to realize experimentally since every energy exchange with the environment has to be suppressed. It is more common to deal with systems exhibiting a fixed temperature T , such as experiments at room temperature. The corresponding ensemble is called *canonical ensemble* and shown in Fig. 1.5.

At a given temperature T , the probability for a system to be in a certain configuration X with energy $E(X)$ is given by

$$\rho(X) = \frac{1}{Z_T} \exp \left[-\frac{E(X)}{k_B T} \right], \quad (1.12)$$

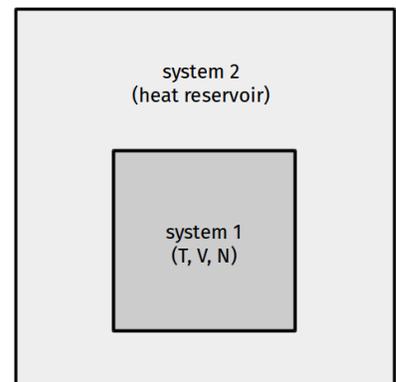


Figure 1.5: In a canonical ensemble setup, the system we study (system 1) is coupled to a heat reservoir (system 2) that guarantees a constant temperature.

with

$$Z_T = \sum_X \exp \left[-\frac{E(X)}{k_B T} \right] \quad (1.13)$$

being the partition function of the canonical ensemble. According to the prior definition in Eq. (1.10), the ensemble average of a quantity Q is then given by

$$\langle Q \rangle = \frac{1}{Z_T} \sum_X Q(X) e^{-\frac{E(X)}{k_B T}}. \quad (1.14)$$

1.3 Ising Model

Due to its historical relevance for the study of phase transitions in statistical physics and its broad applicability in many other fields, we now apply the terminology defined in Sec. 1.2 to the Ising model. Wilhelm Lenz proposed this model to his doctoral student Ernst Ising to describe systems composed of magnetic dipole moments can be in one of two states (+1 or -1). The original goal was to describe phase transitions in magnetic materials. As part of his doctoral thesis in 1924, Ernst Ising showed that the one-dimensional model exhibits no phase transition. For that reason, it was expected that this model was of no particular use. Then in 1949, Lars Onsager published an equation describing the magnetization in the two-dimensional model. Onsager provided no proof for his result, but it is known that he derived it using Toeplitz determinants [5]. A few years passed until a proof has been established by Yang in 1952 [6]. Since then, the Ising model has been successfully applied to a large number of physical and non-physical problems such as magnetic systems, binary mixtures, and models of opinion formation. Also in comparison with experiments, the Ising model has been found to generally agree very well with observations made for certain magnetic materials [7]. To date, no general analytical solution for the Ising model in three dimensions is known. This is the reason why this mathematical model has been studied so intensively from a theoretical and numerical perspective, using tools of statistical physics some of which we describe in this section.

We start our discussion of the properties of the Ising model by considering a two-dimensional lattice with sites $\sigma_i \in \{1, -1\}$ which only interact with their nearest neighbors as shown in Fig. 1.6. This restriction can be relaxed by letting the sites interact with the next nearest neighbors or even farther sites. If we think of the variables as configuration of classical spins $\{\sigma\}$, their interaction is described by the Hamiltonian

$$\mathcal{H}(\{\sigma\}) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - H \sum_{i=1}^N \sigma_i, \quad (1.16)$$

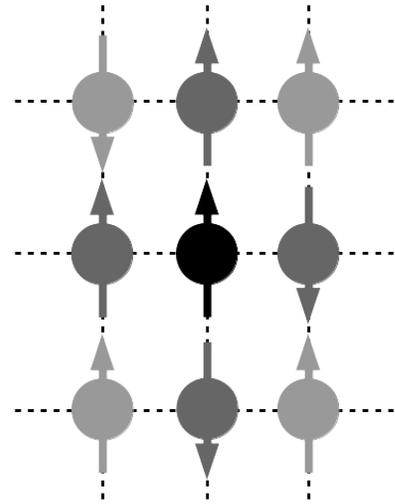
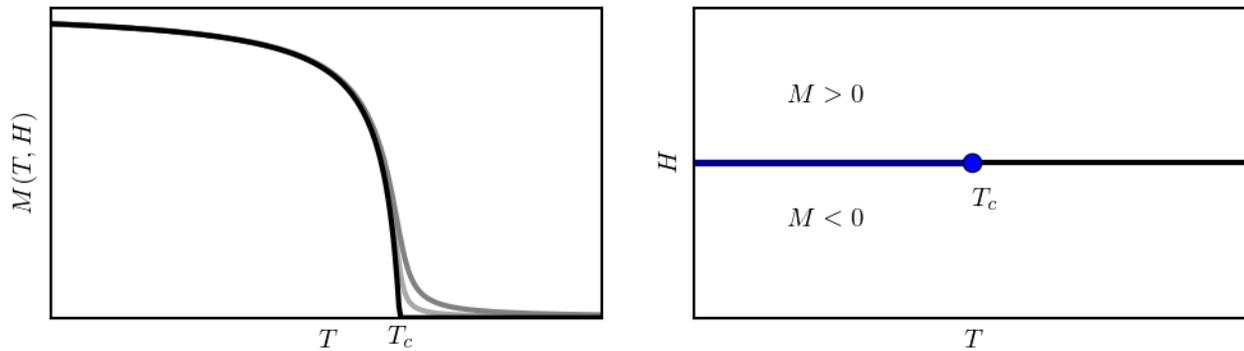


Figure 1.6: An illustration of the interaction of a magnetic dipole (black) with its nearest neighbors (dark grey) on a two-dimensional lattice.

where the first term denotes the interaction between all nearest neighbors represented by a sum over $\langle i, j \rangle$, and the second one the interaction of each site with an external magnetic field H . In the ferromagnetic case where $J > 0$, the energy is lowered ($-J$) if the spins are parallel whereas the energy is increased ($+J$) if the spins are anti-parallel. If $J < 0$, the system is called antiferromagnetic since anti-aligned spins lead to an energy minimization. No interaction occurs for $J = 0$. Considering a ferromagnet, the first term in Eq. (1.16) tries to create order in the system by minimizing the overall energy as a consequence of aligning spins in the same direction. The second term tends to align the spins in the direction of the external field H . While the energy is lower in an ordered state, thermal fluctuations tend to destroy the order by flipping single spins. For temperatures above a



critical value, i.e., $T > T_c$, such fluctuations dominate the system and there is no large-scale alignment of spins observable anymore. However, the domains of aligned spins grow as the temperature decreases below the critical value. This transition, like any other phase transition (e.g., for superconductors, superfluids or sol-gel systems) between ordered and disordered states, can be characterized by an *order parameter*. In the case of the Ising model, the order parameter is the spontaneous magnetization which undergoes a phase transition of second order³ as illustrated in Fig. 1.7. The value of the critical temperature depends on the dimension and the topology of the system. In the case of a two-dimensional square lattice, the critical temperature ($k_B T_c / J \approx 2.269$) has been computed analytically by Onsager [8]. First order transitions also occur in the Ising model as a consequence of a sign change of the external magnetic field. This behavior is also illustrated in Fig. 1.7. Crossing the blue first order transition line at $H = 0$ leads to a jump in the magnetization.

Figure 1.7: In the left panel, we show the magnetization $M(T, H)$ for different fields $H \geq 0$ as a function of T . The black solid line represents the spontaneous magnetization $M_S(T)$ for $H = 0$ and should be interpreted in the sense that $\lim_{H \rightarrow 0^+} M(T, H)$. The right panel illustrates the first-order transition as a consequence of a sign change of the external field.

³ Second order phase transitions are continuous in the order parameter—in the case of the Ising model, the first derivative of the free energy with respect to the magnetic field.

1.3.1 Order Parameter

The magnetization is defined as

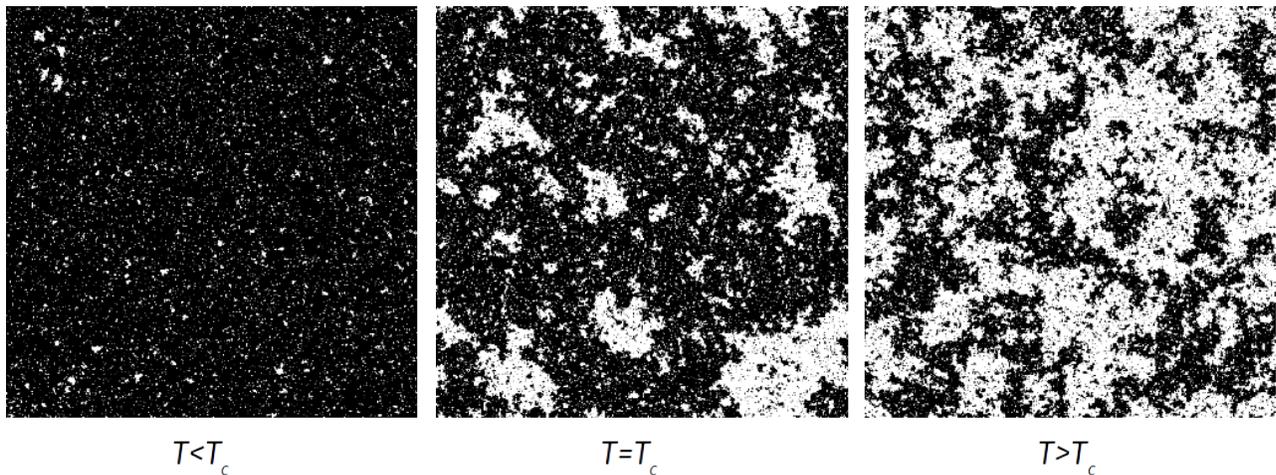
$$M(T, H) = \left\langle \frac{1}{N} \sum_{i=1}^N \sigma_i \right\rangle, \quad (1.17)$$

and corresponds to the ensemble average of the mean value of all spins. In the case of a vanishing external field, the Hamiltonian is invariant under a simultaneous reversal of all spins. In other words, a certain equilibrated configuration of spins would also be of minimal energy if we would change the sign of every single spin. Thus, the ensemble average defined by Eq. (1.17) would not be a good measure of the magnetization since it corresponds to an ensemble average over all possible configurations.

On average, $M(T)$ vanishes since for every configuration there exists one of opposite sign which neutralizes the other one. As a consequence, we define the *order parameter* of the Ising model as

$$M_S(T) = \lim_{H \rightarrow 0^+} \left\langle \frac{1}{N} \sum_{i=1}^N \sigma_i \right\rangle \quad (1.18)$$

and refer to it as the *spontaneous magnetization*. In the definition of $M_S(T)$, the symmetry of the Ising model is broken by applying a vanishing positive field H that aligns the spins in one direction. Another possibility of breaking the symmetry can be realized by keeping the boundaries of the lattice in a certain state. This is, however, impracticable if periodic boundaries are being used. In Fig. 1.8, we illustrate



typical domain configurations for different temperatures. If $T > T_c$, one observes that thermal fluctuations lead to random configurations.

Figure 1.8: The formation of magnetic domains in the Ising model for temperatures $T < T_c$. For larger temperatures, the configurations are random due to thermal fluctuations. The simulations have been performed on a square lattice with 512×512 sites using <https://mattbierbaum.github.io/ising.js/>.

On the other hand, for small enough temperatures magnetic domains start to form. In the vicinity of the critical temperature for $T < T_c$, the spontaneous magnetization scales as

$$M_S(T) \propto (T_c - T)^\beta. \quad (1.19)$$

For $T = T_c$ and $H \rightarrow 0$, we find the following scaling

$$M(T = T_c, H) \propto H^{1/\delta}. \quad (1.20)$$

The exponents β and δ are so-called *critical exponents* and characterize together with other exponents the underlying phase transition.

Different techniques such as series expansions, field-theoretic methods, and very sophisticated Monte Carlo algorithms exist to determine critical exponents and the critical temperature. The Monte Carlo concepts introduced in this course are a common tool to most precisely determine these quantities [9].

1.3.2 Fluctuations

The magnetic susceptibility is defined as the change of the magnetization M in response to an applied magnetic field H , i.e.,

$$\chi(T) = \frac{\partial M(T, H)}{\partial H}. \quad (1.23)$$

We now use the definition of the spontaneous magnetization given by Eq. (1.18) and plug it into Eq. (1.23) leading to

$$\begin{aligned} \chi(T) &= \lim_{H \rightarrow 0^+} \frac{\partial \langle M(T, H) \rangle}{\partial H} \\ &= \lim_{H \rightarrow 0^+} \frac{\partial}{\partial H} \frac{1}{N} \frac{\sum_{\{\sigma\}} \sum_{i=1}^N \sigma_i \exp\left(\frac{E_0 + H \sum_{i=1}^N \sigma_i}{k_B T}\right)}{\underbrace{\sum_{\{\sigma\}} \exp\left(\frac{E_0 + H \sum_{i=1}^N \sigma_i}{k_B T}\right)}_{=Z_T(\mathcal{H})}}. \end{aligned}$$

Here we used the definition of the ensemble average given by Eq. (1.14) with $E_0 = J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$ and the canonical partition function of the Ising Hamiltonian $Z_T(\mathcal{H})$.

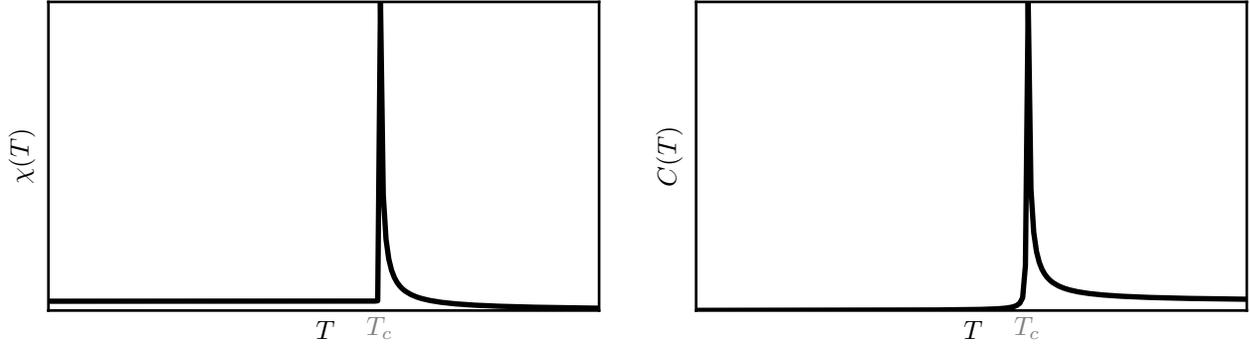
Using the product rule yields

$$\begin{aligned} \chi(T) &= \lim_{H \rightarrow 0^+} \frac{1}{N k_B T} \frac{\sum_{\{\sigma\}} \left(\sum_{i=1}^N \sigma_i\right)^2 \exp\left(\frac{E_0 + H \sum_{i=1}^N \sigma_i}{k_B T}\right)}{Z_T(\mathcal{H})} - \frac{1}{N k_B T} \frac{\left[\sum_{\{\sigma\}} \sum_{i=1}^N \sigma_i \exp\left(\frac{E_0 + H \sum_{i=1}^N \sigma_i}{k_B T}\right)\right]^2}{[Z_T(\mathcal{H})]^2} \\ &= \frac{N}{k_B T} \langle M_S(T)^2 \rangle - \frac{N}{k_B T} \langle M_S(T) \rangle^2 \\ &= \frac{N}{k_B T} \left[\langle M_S(T)^2 \rangle - \langle M_S(T) \rangle^2 \right] \geq 0. \end{aligned} \quad (1.26)$$

The last equation defines fluctuation-dissipation theorem for the magnetic susceptibility. Analogously, the specific heat is connected to energy fluctuations since

$$C(T) = \frac{\partial \langle E \rangle}{\partial T} = \frac{1}{k_B T^2} \left[\langle E(T)^2 \rangle - \langle E(T) \rangle^2 \right]. \quad (1.27)$$

Eqs. (1.26) and (1.27) describe the variance of magnetization and energy, respectively. Similarly to the power-law scaling of the spon-



taneous magnetization defined in Eq. (1.19), we find for the magnetic susceptibility in the vicinity of T_c

$$\chi(T) \propto |T_c - T|^{-\gamma} \quad (1.28)$$

with $\gamma = 7/4$ in two dimensions and approximately 1.24 in three dimensions. Moreover, the specific heat exhibits the following scaling

$$C(T) \propto |T_c - T|^{-\alpha}, \quad (1.29)$$

where $\alpha = 0$ in two dimensions⁴ and numerically known in three dimensions with $\alpha \approx 0.11$. The temperature dependence of susceptibility and specific heat is illustrated in Fig. 1.9.

1.3.3 Correlation Length

The correlation function is defined by

$$G(r_1, r_2; T, H) = \langle \sigma_1 \sigma_2 \rangle - \langle \sigma_1 \rangle \langle \sigma_2 \rangle, \quad (1.32)$$

where the vectors r_1 and r_2 pointing in the direction of lattice sites 1 and 2. If the system is translational and rotational invariant, the correlation function only depends on $r = |r_1 - r_2|$. At the critical point, the correlation function decays as

$$G(r; T_c, 0) \propto r^{-d+2-\eta}, \quad (1.33)$$

Figure 1.9: Susceptibility and specific heat as a function of temperature for the three dimensional Ising model. Both quantities diverge at the critical temperature T_c in the thermodynamic limit.

⁴ An exponent of $\alpha = 0$ corresponds to a logarithmic decay since

$$\lim_{s \rightarrow 0} \frac{|x|^{-s} - 1}{s} = -\ln |x|.$$

However, in many cases it is difficult to decide if the exponent is zero or just has a small value [4].

where η is another critical exponent and d the dimension of the system. In two and three dimensions, the value of η is given by $1/4$ and 0.036 , respectively. For temperatures away from the critical temperature, the correlation function exhibits an exponential decay

$$G(r; T, 0) \propto r^{-\vartheta} e^{-r/\xi}, \quad (1.34)$$

where ξ defines the *correlation length*. The exponent ϑ equals 2 above and $1/2$ below the transition point. In the vicinity of T_c , the correlation length ξ diverges since

$$\xi(T) \propto |T_c - T|^{-\nu} \quad (1.35)$$

with $\nu = 1$ in two dimensions and $\nu \approx 0.63$ in three dimensions.

1.3.4 Critical Exponents and Universality

The aforementioned six critical exponents are connected by four scaling laws

$$\alpha + 2\beta + \gamma = 2 \quad (\text{Rushbrooke}), \quad (1.41)$$

$$\gamma = \beta(\delta - 1) \quad (\text{Widom}), \quad (1.42)$$

$$\gamma = (2 - \eta)\nu \quad (\text{Fisher}), \quad (1.43)$$

$$2 - \alpha = d\nu \quad (\text{Josephson}), \quad (1.44)$$

which have been derived in the context of the phenomenological scaling theory for ferromagnetic systems [10, 11]. Due to these relations, the number of independent exponents reduces to two. The Josephson law includes the spatial dimension d of the system and thus defines a *hyperscaling* relation. Such relations are valid only below the upper critical dimension which is equal to $d_c = 4$ for the Ising model.

The importance of critical exponents becomes more clear when we study different systems exhibiting phase transitions. Critical control parameters such as T_c in the Ising model sensitively depend on the interaction details. However, critical exponents only depend on fundamental system properties such as dimension or symmetries and are therefore said to be *universal*. Based on these observations, the universality hypothesis states that different critical phenomena can be reduced to a small number of *universality classes* [13]. All systems belonging to a certain universality class share the same critical exponents and the same scaling behavior near the critical point.

Universality is not a mere theoretical concept but can also be observed experimentally. In Fig. 1.10, we show an example of five different fluids undergoing a liquid-gas transition. All substances exhibit different inter-atomic interactions, and still we observe a clear data collapse for the rescaled chemical potential.

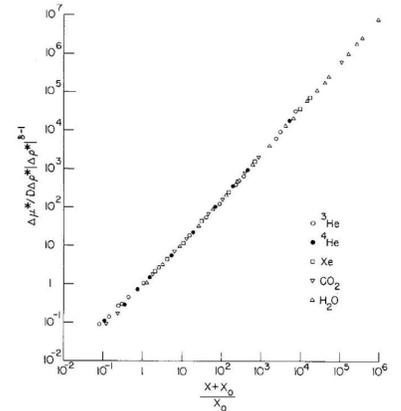


Figure 1.10: Universal scaling for five different gases. The scaling variable is defined as $x = \Delta T |\Delta\rho|^{-1/\beta}$ and x_0 depends on the amplitude B of the power-law for the coexistence curve $\Delta\rho = B\Delta T^\beta$ [4]. The figure is taken from Ref. [12].

Table 1.1: The critical exponents of the Ising model in two and three dimensions [14].

Exponent	$d = 2$	$d = 3$
α	0	0.110(1)
β	$1/8$	0.3265(3)
γ	$7/4$	1.2372(5)
δ	15	4.789(2)
η	$1/4$	0.0364(5)
ν	1	0.6301(4)

1.4 Monte Carlo Methods

Monte Carlo methods are broadly applicable to different problems which are (i) difficult (or impossible) to solve analytically, or (ii) not solvable with other numerical techniques due to the large computational complexity. For example, it may be difficult to numerically sample the complete phase or parameter space of a certain model. Instead, we could apply an appropriate random sampling technique to explore its properties.

In this section, we focus on Monte Carlo algorithms to numerically study the properties of the Ising model and other Hamiltonian systems. The basic idea behind the concepts we develop is that randomly sampling the phase space instead of averaging over all states is sufficient to compute the ensemble average of a certain thermodynamic quantity. If the number of samples is large enough, the computed estimate eventually converges towards the real value.

The main steps of the Monte Carlo sampling are

1. Choose randomly a new configuration in phase space based on a Markov chain.
2. Accept or reject the new configuration, depending on the strategy used (e.g., Metropolis or Glauber dynamics).
3. Compute the physical quantity and add it to the averaging procedure.
4. Repeat the previous steps.

1.4.1 Markov Chains

In most cases, a sampling based on equally distributed configurations is very inefficient since the underlying distribution may exhibit peaks or another form which is not uniform. As an example, we consider the kinetic energy of an ideal gas. The distribution of the mean energy will be a sharp peak as depicted in Fig. 1.11. There exist many different methods which avoid unnecessary sampling of regions where the system is unlikely to be found (*importance sampling*). A common way to efficiently choose appropriate samples out of the large pool of possible configurations, is to explore the phase space using a *Markov chain*. We therefore introduce the virtual time τ and note that it only represents the steps of a stochastic process and should not be confused with physical time.

In terms of a Markov chain, the transition probability from one state to another is given by the probability of a new state to be proposed (T) and the probability of this state to be accepted (A). Specifically,

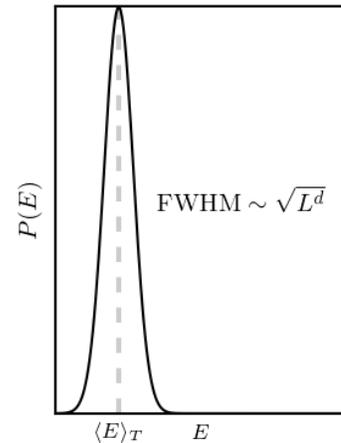


Figure 1.11: Example of an energy distribution with a system size L dependence of the distribution width which scales as $\propto \sqrt{L^d}$ where d is the system dimension.

$T(X \rightarrow Y)$ is the probability that a new configuration Y is proposed, starting from configuration X . For the thermodynamic systems we consider, the transition probability fulfills three conditions:

1. *Ergodicity*: any configuration in the phase space must be reachable within a finite number of steps,
2. *Normalization*: $\sum_Y T(X \rightarrow Y) = 1$,
3. *Reversibility*: $T(X \rightarrow Y) = T(Y \rightarrow X)$.

Once a configuration is proposed, we can accept the new configuration with probability $A(X \rightarrow Y)$ or reject it with probability $1 - A(X \rightarrow Y)$. The *probability of the Markov chain* is then given by

$$W(X \rightarrow Y) = T(X \rightarrow Y) \cdot A(X \rightarrow Y). \quad (1.45)$$

We denote the probability to find the system in a certain configuration X at virtual time τ by $p(X, \tau)$. The *master equation* describes the time evolution of $p(X, \tau)$ and is given by

$$\frac{dp(X, \tau)}{d\tau} = \sum_Y p(Y)W(Y \rightarrow X) - \sum_Y p(X)W(X \rightarrow Y). \quad (1.46)$$

A stationary state p_{st} is reached if $\frac{dp(X, \tau)}{d\tau} = 0$. The probability of the Markov chain fulfills the following properties:

1. *Ergodicity*: any configuration must be reachable: $\forall X, Y : W(X \rightarrow Y) \geq 0$,
2. *Normalization*: $\sum_Y W(X \rightarrow Y) = 1$,
3. *Homogeneity*: $\sum_Y p_{\text{st}}(Y)W(Y \rightarrow X) = p_{\text{st}}(X)$.

Note that the condition of reversibility is not required anymore. To sample the phase space of our system, we have to come up with an expression for the Markov chain probability $W(\cdot)$. As an example, we consider a two level system in which one of the two energy levels is higher (e.g., the electronic shells in an atom): At low energies it would be unnecessary to equally sample the excited and the ground state of the electrons. On the contrary, at very high energies the sampling will have to reflect the higher probability of an electron to be in an excited state, rather than in the ground state.

To efficiently sample the relevant regions of the phase space, the probability of the Markov chain $W(\cdot)$ has to depend on the system properties. To achieve that, we set the stationary distribution p_{st} equal to the equilibrium distribution of the physical system p_{eq} (a real and measurable distribution):

$$\frac{dp(X, \tau)}{d\tau} = 0 \Leftrightarrow p_{\text{st}} \stackrel{!}{=} p_{\text{eq}}. \quad (1.47)$$

It then follows from the stationary state condition of the Markov chain that

$$\sum_Y p_{\text{eq}}(Y)W(Y \rightarrow X) = \sum_Y p_{\text{eq}}(X)W(X \rightarrow Y). \quad (1.48)$$

A sufficient condition for this to be true is

$$p_{\text{eq}}(Y)W(Y \rightarrow X) = p_{\text{eq}}(X)W(X \rightarrow Y), \quad (1.49)$$

which is referred to as *condition of detailed balance*.

As an example, in a canonical ensemble at fixed temperature T , the equilibrium distribution is given by the Boltzmann distribution

$$p_{\text{eq}}(X) = \frac{1}{Z_T} \exp \left[-\frac{E(X)}{k_B T} \right] \quad (1.50)$$

with the partition function $Z_T = \sum_X \exp \left[-\frac{E(X)}{k_B T} \right]$.

1.4.2 $M(RT)^2$ Algorithm

One possible choice of the acceptance probability fulfilling the detailed balance condition is given by

$$A(X \rightarrow Y) = \min \left[1, \frac{p_{\text{eq}}(Y)}{p_{\text{eq}}(X)} \right]. \quad (1.51)$$

which can be obtained by rewriting Eq. (1.49). This algorithm has been developed at Los Alamos National Laboratory in the group of Nicolas Metropolis whose portrait is shown in Fig. 1.12. It is often referred to as Metropolis or $M(RT)^2$ algorithm⁵. In the case of the canonical ensemble with $p_{\text{eq}}(X) = \frac{1}{Z_T} \exp \left[-\frac{E(X)}{k_B T} \right]$, the acceptance probability becomes

$$A(X \rightarrow Y) = \min \left[1, \exp \left(-\frac{\Delta E}{k_B T} \right) \right], \quad (1.52)$$

where $\Delta E = E(Y) - E(X)$. The last equation implies that the Monte Carlo step is always accepted if the energy decreases, and if the energy increases, it is accepted with probability $\exp \left(-\frac{\Delta E}{k_B T} \right)$. Plugging Eq. (1.51) with the Boltzmann factor p_{eq} into Eq. (1.49) shows that detailed balance is indeed fulfilled. The algorithm has been generalized in 1970 [19]. We use the rather general algorithm to explore the phase space of the Ising model by flipping the values on the magnetic monopoles according to the acceptance probability as defined by Eq. (1.52). In Fig. 1.13, we show the spontaneous magnetization $M_S(T)$ and the energy per spin $E(T)$ of the three-dimensional Ising model simulated with the $M(RT)^2$ algorithm.

In summary, the steps of the $M(RT)^2$ algorithm applied to the Ising model are

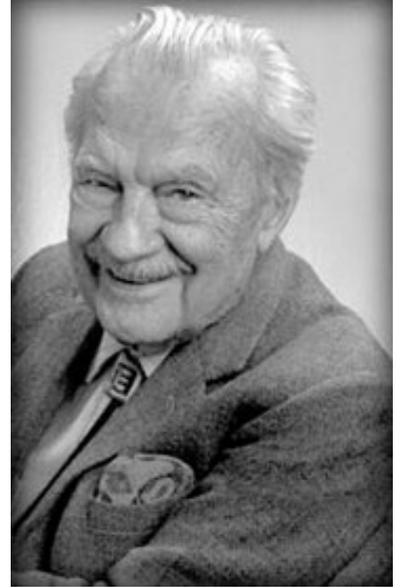


Figure 1.12: Nicholas C. Metropolis (1915-1999) introduced the $M(RT)^2$ sampling technique.

$M(RT)^2$ is an abbreviation of the last names of the authors of the original paper [15]. RT is squared because except Metropolis, the other four authors of the paper formed two married couples and therefore carried the same family names. The real contributions of some of the authors (in particular of Metropolis and of A.H. Teller) is subject of controversy [16, 17]. It has been even stated by Roy Glauber and Emilio Segré that the original algorithm was invented by Enrico Fermi [18].

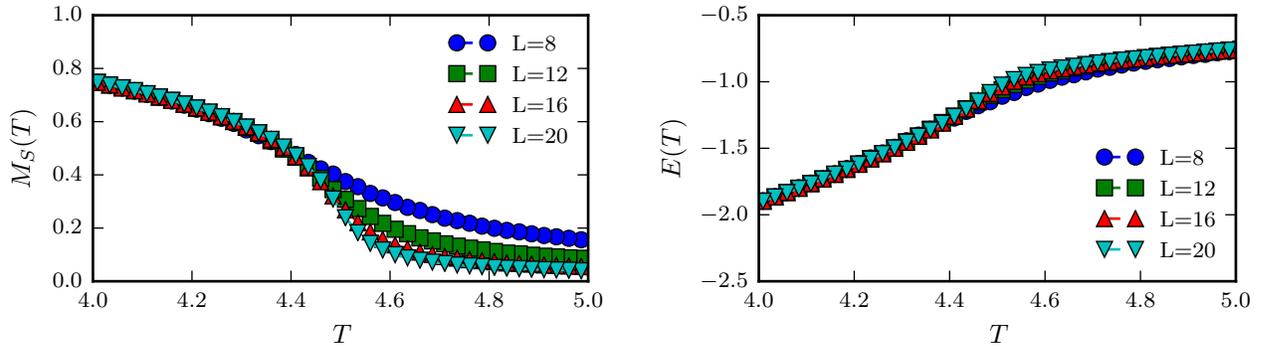


Figure 1.13: We show the spontaneous magnetization $M_S(T)$ and the energy per spin $E(T)$ of the three-dimensional Ising model for different temperatures T and linear system sizes L .

M(RT)² Algorithm

- Randomly choose a lattice site i ,
- Compute $\Delta E = E(Y) - E(X) = 2J\sigma_i h_i$,
- Flip the spin if $\Delta E \leq 0$, otherwise accept it with probability $\exp\left(-\frac{\Delta E}{k_B T}\right)$,

with $h_i = \sum_{\langle i,j \rangle} \sigma_j$ and $E = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$. One should bear in mind that there is only a limited number of possible energy differences in a lattice. To speed up the simulation, we can therefore create a lookup table storing possible nearest neighbor combinations. For example $h_i \in \{0, \pm 2, \pm 4\}$ in the case of a two dimensional lattice.

1.4.3 Glauber Dynamics (Heat Bath Dynamics)

The Metropolis algorithm is not the only possible choice to fulfill the detailed balance condition. Another acceptance probability given by

$$A_G(X \rightarrow Y) = \frac{\exp\left(-\frac{\Delta E}{k_B T}\right)}{1 + \exp\left(-\frac{\Delta E}{k_B T}\right)} \quad (1.53)$$

has been suggested by Glauber.

In contrast to the M(RT)² acceptance probability, updates with $\Delta E = 0$ are not always accepted but with probability 1/2. This behavior is illustrated in Fig. 1.14.

To proof that Eq. (1.53) satisfies the condition of detailed balance, we have to show that

$$p_{\text{eq}}(Y)A_G(Y \rightarrow X) = p_{\text{eq}}(X)A_G(X \rightarrow Y), \quad (1.54)$$

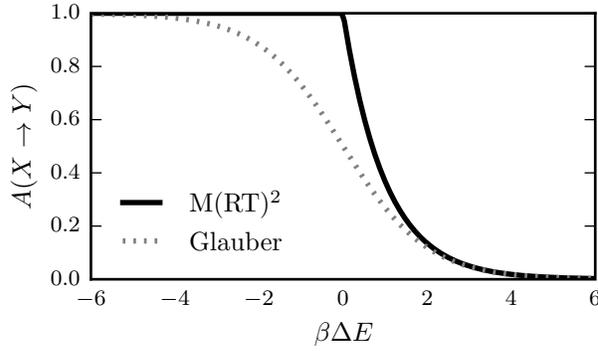


Figure 1.14: A comparison of the acceptance probabilities of $M(RT)^2$ and Glauber dynamics.

since $T(Y \rightarrow X) = T(X \rightarrow Y)$. The last equation is equivalent to

$$\frac{p_{\text{eq}}(Y)}{p_{\text{eq}}(X)} = \frac{A_G(X \rightarrow Y)}{A_G(Y \rightarrow X)} \quad (1.55)$$

which is fulfilled since

$$\frac{p_{\text{eq}}(Y)}{p_{\text{eq}}(X)} = \exp\left(-\frac{\Delta E}{k_B T}\right) \quad (1.56)$$

and

$$\frac{A_G(X \rightarrow Y)}{A_G(Y \rightarrow X)} = \frac{\exp\left(-\frac{\Delta E}{k_B T}\right)}{1 + \exp\left(-\frac{\Delta E}{k_B T}\right)} \left[\frac{\exp\left(\frac{\Delta E}{k_B T}\right)}{1 + \exp\left(\frac{\Delta E}{k_B T}\right)} \right]^{-1} = \exp\left(-\frac{\Delta E}{k_B T}\right). \quad (1.57)$$

As in the $M(RT)^2$ algorithm, only the local configuration around the lattice site is relevant for the update procedure. Furthermore, with $J = 1$, the probability to flip spin σ_i is

$$A_G(X \rightarrow Y) = \frac{\exp\left(\frac{-2\sigma_i h_i}{k_B T}\right)}{1 + \exp\left(\frac{-2\sigma_i h_i}{k_B T}\right)} \quad (1.58)$$

with $h_i = \sum_{(i,j)} \sigma_j$ being the local field and $X = \{\dots, \sigma_{i-1}, \sigma_i, \sigma_{i+1}, \dots\}$ and $Y = \{\dots, \sigma_{i-1}, -\sigma_i, \sigma_{i+1}, \dots\}$ the initial and final configuration, respectively. We abbreviate the probability defined by Eq. (1.58) when the spin at position i point down as p_i . The spin flip and no flip probabilities can then be expressed as

$$p_{\text{flip}} = \begin{cases} p_i & \text{for } \sigma_i = -1 \\ 1 - p_i & \text{for } \sigma_i = +1 \end{cases} \quad \text{and} \quad p_{\text{no flip}} = \begin{cases} 1 - p_i & \text{for } \sigma_i = -1 \\ p_i & \text{for } \sigma_i = +1 \end{cases} \quad (1.59)$$

A possible implementation is

$$\sigma_i(\tau + 1) = -\sigma_i(\tau) \cdot \text{sign}(p_i - z), \quad (1.60)$$

with $z \in (0, 1)$ being a uniformly distributed random number, or

$$\sigma_i(\tau + 1) = \begin{cases} +1 & \text{with propability } p_i \\ -1 & \text{with propability } 1 - p_i \end{cases} \quad \text{and} \quad p_i = \frac{\exp(2\beta h_i)}{1 + \exp(2\beta h_i)}. \quad (1.61)$$

This method does not depend on the spin value at time t and is called *heat bath Monte Carlo*.

1.4.4 Binary Mixtures and Kawasaki Dynamics

We now consider a system of two species A and B (spin up and spin down particles, two different gas molecules, etc.) which are distributed with given concentrations on the sites of a lattice. For example, these two species could represent two sorts of metallic atoms whose numbers are conserved. An illustration of such a system is shown in Fig. 1.15. We model such a mixture by defining

- E_{AA} as the energy of an $A - A$ bond,
- E_{BB} as the energy of a $B - B$ bond,
- E_{AB} as the energy of an $A - B$ bond.

We set $E_{AA} = E_{BB} = 0$ and $E_{AB} = 1$. The Kawasaki dynamics corresponds to a $M(RT)^2$ or Glauber algorithm with constant numbers of spins in each population. The update algorithm is as follows:

Kawasaki dynamics

- Choose a $A - B$ bond,
- Compute ΔE for $A - B \rightarrow B - A$,
- Metropolis: If $\Delta E \leq 0$ flip, else flip with probability

$$p = \exp\left(\frac{-\Delta E}{k_B T}\right),$$

- Glauber: Flip with probability

$$p = \exp\left(-\frac{\Delta E}{k_B T}\right) / \left[1 + \exp\left(-\frac{\Delta E}{k_B T}\right)\right].$$

This procedure is very similar to the previously discussed update schemes. The only difference is that the magnetization is kept constant.

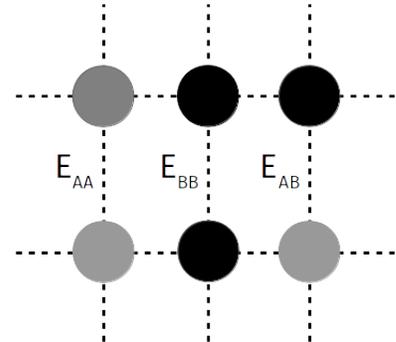


Figure 1.15: An example of a binary mixture consisting of two different atoms A and B .

1.4.5 Creutz Algorithm

Until now, we only considered canonical Monte Carlo algorithms for simulating systems at constant temperature. In 1983, Creutz sug-

gested a method to perform microcanonical Monte Carlo simulations of systems at constant energy [20]. The algorithm is often referred to as *Creutz* algorithm. In fact, the condition of energy conservation is slightly relaxed in this algorithm and energy is not fully conserved. The movement in phase space is therefore not strictly constrained to a subspace of constant energy but there is a certain additional volume in which we can freely move. The condition of constant energy is softened by introducing a so-called *demon* which corresponds to a small reservoir of energy E_D that can store a certain maximum energy E_{\max} .

The Creutz algorithm is defined by the following steps:

Creutz Algorithm

- Choose a site,
- Compute ΔE for the spin flip,
- Accept the change if $E_{\max} \geq E_D - \Delta E \geq 0$.

Besides the fact that we can randomly choose a site, this method involves no random numbers and is thus said to be *deterministic* and therefore *reversible*. The drawback of this method is that the temperature of the system is not known.

It is, however, possible to estimate the temperature with the Boltzmann distribution. By plotting a histogram of the energies E_D one observes a distribution $P(E_D) \propto \exp\left(-\frac{E_D}{k_B T}\right)$. We show an example of a distribution $P(E_D)$ in Fig. 1.16. The fit is not optimal due to the small number of different values of E_D . The larger E_{\max} , the faster the method since the condition of constant energy is relaxed and the exploration of phase space is less restricted to certain regions.

1.4.6 Q2R

In the case of $E_{\max} \rightarrow 0$, the Creutz algorithm resembles a totalistic cellular automaton called *Q2R* [21]. The update rules on a square lattice for spins $\sigma_{ij} \in \{0, 1\}$ are given by

$$\sigma_{ij}(\tau + 1) = f(x_{ij}) \oplus \sigma_{ij}(\tau) \quad (1.62)$$

with

$$x_{ij} = \sigma_{i-1j} + \sigma_{i+1j} + \sigma_{ij-1} + \sigma_{ij+1} \quad \text{and} \quad f(x) = \begin{cases} 1 & \text{if } x = 2 \\ 0 & \text{if } x \neq 2 \end{cases}. \quad (1.63)$$

In this case, the spins are flipped if and only if the change in energy is zero. This can be implemented in a very efficient way using multi spin coding, cf. Sec. 1.10.1. The update rule of Eq. (1.62) can be expressed

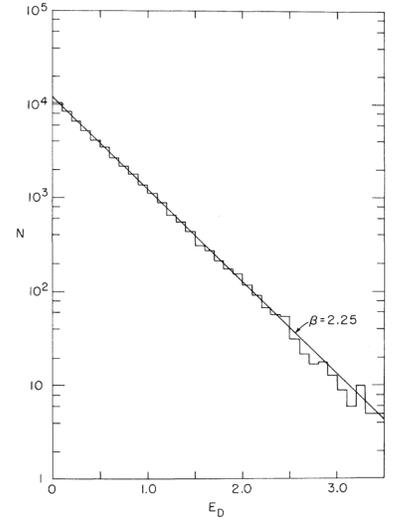


Figure 1.16: The distribution of the demon energy E_D is exponentially distributed. Based on the Boltzmann factor, it is possible to extract the inverse temperature $\beta = (k_B T)^{-1} = 2.25$. The figure is taken from Ref. [20].

in a very elegant way using logical functions

$$\sigma(\tau + 1) = \sigma(\tau) \oplus \{[(\sigma_1 \oplus \sigma_2) \wedge (\sigma_3 \oplus \sigma_4)] \vee [(\sigma_1 \oplus \sigma_3) \wedge (\sigma_2 \oplus \sigma_4)]\}. \quad (1.64)$$

This logical expression can be computed in roughly 12 cycles which last typically around 10 ns. The method is extremely fast, deterministic and reversible. The problem is that it is not ergodic and that it strongly depends on the initial configuration. As an example, try to imagine the evolution of a small lattice, in which only $\sigma_{2,1}$ and $\sigma_{1,2}$ are equal to unity. In fact, this method is not used in statistical physics but it is useful for other purposes, e.g., neuroinformatics or cellular automata.

1.4.7 Boundary Conditions

Computer simulations are always restricted to finite domains. Therefore, one of the finite size effects one has to take into account is the effect of boundaries of the considered domain. The values there can be set to a certain value or periodic boundaries can be introduced. In some situations certain boundary condition also correspond to a real physical situation. As an example, we can think of some zero potential boundary condition while solving the Laplace equation in electrostatics using finite difference methods. There are more effects related to systems of finite size which we describe in more detail in Sec. 1.5. For finite lattices, the following boundary conditions may be used:

- Open boundaries, i.e., no neighbors at the edges of the system,
- fixed boundary conditions,
- and periodic boundaries.

If our system is large enough⁶, it is possible to connect the edges of the lattice. Identifying the last element of a row (or column) with the first element of the next row (or column) leads to so called *helical boundary condition*.

⁶ In this context, *large* is not a purely arbitrary concept but it depends on what we want to simulate. A good measure of *large* can be that the edges of the system are uncorrelated. It is clear that this method is useless in *small* lattices.

1.4.8 Temporal Correlations

Each time we accept a spin-flip in our sampling chain, a new configuration is generated. The problem is that both new and previous configurations are strongly correlated. As a consequence, we can expect that the error scaling of our Monte Carlo method with $\propto 1/\sqrt{N}$, where N is the number of sampled configurations, is no longer valid. We thus have to find a measure which tells us whether we are already in equilibrium or not, and which enables us to sample uncorrelated

configurations. According to the definition of a Markov chain, the dependence of a quantity A on virtual time τ is given by

$$\langle A(\tau) \rangle = \sum_X p(X, \tau) A(X) = \sum_X p(X, \tau_0) A(X(\tau)). \quad (1.65)$$

In the second step of the last equation, we used the fact that the average is taken over an ensemble of initial configurations $X(\tau_0)$ which evolve according to Eq. (1.46) [22]. For some $\tau_0 < \tau$, the *non-linear correlation function*

$$\Phi_A^{\text{nl}}(\tau) = \frac{\langle A(\tau) \rangle - \langle A(\infty) \rangle}{\langle A(\tau_0) \rangle - \langle A(\infty) \rangle} \quad (1.66)$$

is a measure to quantify the deviation of $A(\tau)$ from $A(\infty)$ relative to the deviation of $A(\tau_0)$ from $A(\infty)$. In a simulation, we would consider a large value of τ to define $A(\infty)$. If $A(\tau_0)$ is already equilibrated, we would expect that $\Phi_A^{\text{nl}}(\tau) = 1$. Strictly speaking, Eq. (1.66) is not a correlation function, but it can be a measure to investigate the correlation of configurations.

The *non-linear correlation time* τ_A^{nl} describes the relaxation towards equilibrium and is defined as⁷

$$\tau_A^{\text{nl}} = \int_0^{\infty} \Phi_A^{\text{nl}}(\tau) d\tau. \quad (1.68)$$

In the vicinity of the critical temperature T_c , we observe the so-called *critical slowing down* of our dynamics, i.e., the non-linear correlation time is described by power law

$$\tau_A^{\text{nl}} \propto |T - T_c|^{-z_A^{\text{nl}}} \quad (1.69)$$

with z_A^{nl} being the non-linear dynamical critical exponent. This is very bad news, because the last equation implies that the time needed to reach equilibrium diverges at T_c . The linear correlation function of two quantities A and B in equilibrium is defined as

$$\Phi_{AB}(\tau) = \frac{\langle A(\tau_0)B(\tau) \rangle - \langle A \rangle \langle B \rangle}{\langle AB \rangle - \langle A \rangle \langle B \rangle} \quad (1.70)$$

with

$$\langle A(\tau_0)B(\tau) \rangle = \sum_X p(X, \tau_0) A(X(\tau_0)) B(X(\tau)).$$

As τ goes to infinity, $\Phi_{AB}(\tau)$ decreases from unity to zero. If $A = B$, we call Eq. (1.70) the *autocorrelation function*. For the spin-spin correlation in the Ising model we obtain

$$\Phi_{\sigma}(\tau) = \frac{\langle \sigma(\tau_0)\sigma(\tau) \rangle - \langle \sigma(\tau_0) \rangle^2}{\langle \sigma^2(\tau_0) \rangle - \langle \sigma(\tau_0) \rangle^2}. \quad (1.71)$$

⁷ If we consider an exponential decay of $\Phi_A^{\text{nl}}(\tau)$, we find that this definition is meaningful since

$$\int_0^{\infty} \exp(-\tau/\tau_A^{\text{nl}}) d\tau = \tau_A^{\text{nl}}. \quad (1.67)$$

The *linear* correlation time τ_{AB} describes the relaxation in equilibrium

$$\tau_{AB} = \int_0^{\infty} \Phi_{AB}(\tau) d\tau. \quad (1.72)$$

As in the case of the non-linear correlation time, in the vicinity of T_c , we observe a *critical slowing down*, i.e.,

$$\tau_{AB} \propto |T - T_c|^{-z_A}. \quad (1.73)$$

with z_A being the *linear* dynamical critical exponent.

The dynamical exponents for spin correlations turn out to be

$$z_\sigma = 2.16 \text{ (2D)}, \quad (1.74)$$

$$z_\sigma = 2.09 \text{ (3D)}. \quad (1.75)$$

There is a conjectured relation between the Ising critical exponents and the critical dynamical exponents for spin σ and energy correlations E . The relations

$$z_\sigma - z_\sigma^{\text{nl}} = \beta, \quad (1.76)$$

$$z_E - z_E^{\text{nl}} = 1 - \alpha, \quad (1.77)$$

are numerically well-established, however, not yet analytically proven.

1.4.9 Decorrelated Configurations

The correlation behavior described in the previous Sec. 1.4.8 is only valid in the case of an infinite lattice. The correlation length diverges at T_c according to Eq. (1.35). In a finite system, however, we cannot observe a quantity diverging towards infinity—the correlation length ξ approaches the system size L at T_c . This behavior is illustrated in Fig. 1.17. Connecting this behavior with the one observed for the correlation time described by Eq. (1.72) yields

$$\tau_{AB} \propto |T - T_c|^{-z_{AB}} \propto L^{\frac{z_{AB}}{\nu}}. \quad (1.78)$$

The last equation implies that the number of samples which have to be discarded increases with system size. This is a problem when studying large system sizes, because the computation may take very long. To ensure not to sample correlated configurations one should

- first reach equilibrium (discard $n_0 = c\tau^{\text{nl}}(T)$ configurations),
- only sample every $n_c^{\text{th}} = c\tau(T)$ configuration,

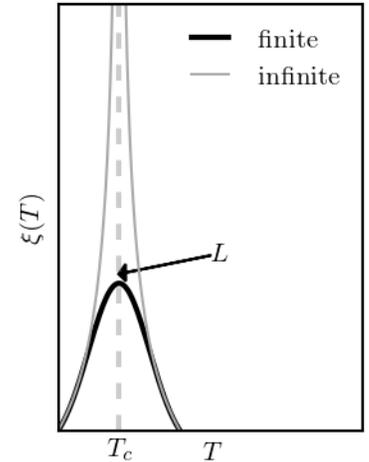


Figure 1.17: The correlation length diverges in an infinite system at T_c according to Eq. (1.35). In a finite system, however, we observe a round off and the correlation length approaches the system size L at T_c .

- and at T_c use $n_0 = cL^{\frac{z\eta_l}{\nu}}$ and $n_e = cL^{\frac{z}{\nu}}$

where $c \approx 3$ is a "safety factor" to make sure to discard enough samples. A trick for reducing this effect is using cluster algorithms which we introduce in Sec. 1.6.

1.5 Finite Size Methods

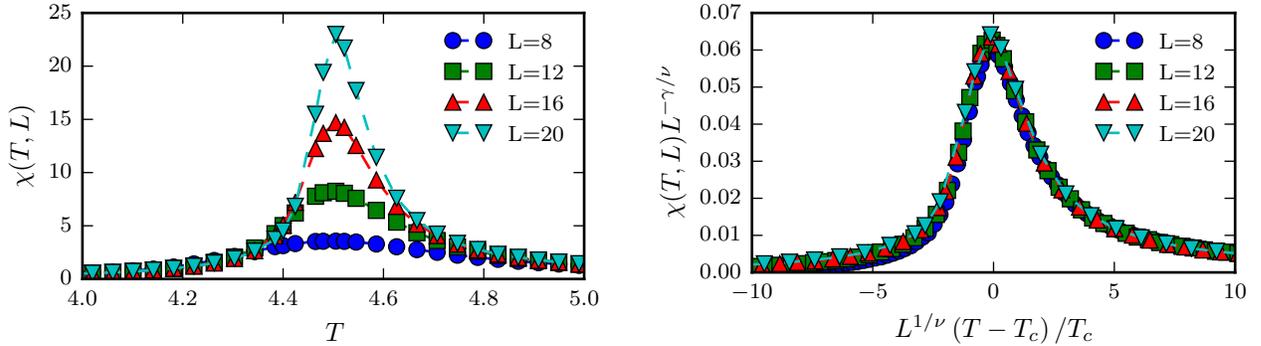


Figure 1.18: The system size dependence of the susceptibility and the corresponding finite size scaling. The data is based on a simulation of the Ising model on a cubic lattice for different linear system sizes L . The number of samples is 25×10^3 which have been generated with the Wolf algorithm. The error bars are smaller than the markers.

In the Ising model, spins tend to form clusters and spatial correlations emerge in addition to the temporal correlations discussed in Sec. 1.4.8. In particular, we introduced the concepts of correlation function and correlation length in Sec. 1.3.3. We described that the correlation function decays exponentially with distance and that the correlation length (finite system size) defines a characteristic length scale. According to Eq. (1.35), the correlation length diverges with an exponent of $-\nu$ at T_c . Moreover, susceptibility and heat capacitance also exhibit divergent behavior at T_c as described by Eqs. (1.28) and (1.29). The larger the system size, the more pronounced is the divergence.

In finite systems, we cannot observe a divergence towards infinity but a peak of some finite value. The value of the susceptibility peak scales with $L^{\frac{\gamma}{\nu}}$ while the critical region shrinks as $L^{-\frac{1}{\nu}}$. If we rescale the values for different system sizes accordingly, we obtain a data collapse, i.e., all the values fall onto a single curve. This can be used to compute critical exponents.

The finite size scaling relation of the susceptibility is given by

$$\chi(T, L) = L^{\frac{\gamma}{\nu}} F_\chi \left[(T - T_c) L^{\frac{1}{\nu}} \right], \quad (1.80)$$

where F_χ is called susceptibility *scaling function*⁸. We see an example

⁸ Based on Eq. (1.28), we can infer that $F_\chi \left[(T - T_c) L^{\frac{1}{\nu}} \right] \propto \left(|T - T_c| L^{\frac{1}{\nu}} \right)^{-\gamma}$ as $L \rightarrow \infty$.

of a finite size scaling data collapse of the susceptibility in Fig. 1.18.

In the case of the magnetization, the corresponding finite size scaling relation is

$$M_S(T, L) = L^{-\frac{\beta}{\nu}} F_{M_S} \left[(T - T_c) L^{\frac{1}{\nu}} \right]. \quad (1.81)$$

1.5.1 Binder Cumulant

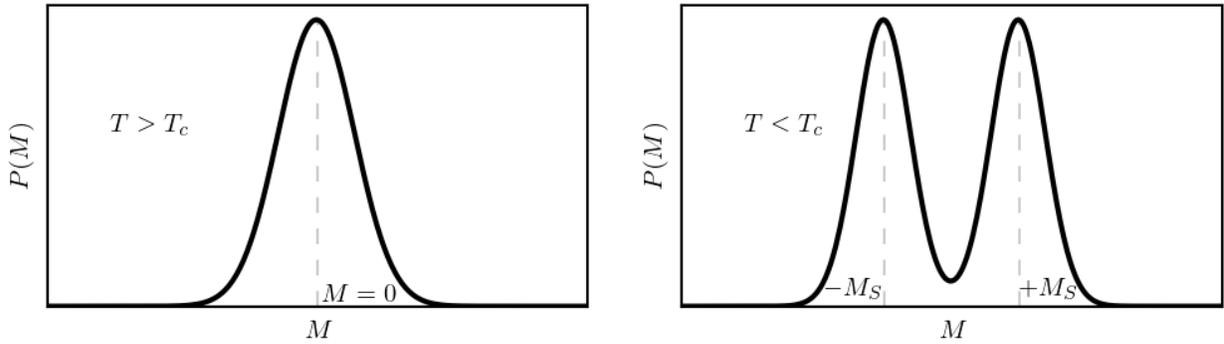


Figure 1.19: The distribution $P(M)$ of the magnetization M above a below the critical temperature T_c .

In the previous section, we have outlined that the finite size scaling relation of the susceptibility according to Eq. (1.80) constitutes a method to determine critical exponents. However, we still need a way to determine the critical temperature more precisely than with a mere observation of the initial growth of the spontaneous magnetization as shown in Fig. 1.7. To overcome this hurdle, we make use of the so-called *Binder cumulant*

$$U_L(T) = 1 - \frac{\langle M^4 \rangle_L}{3 \langle M^2 \rangle_L^2}, \quad (1.84)$$

which is independent of the system size L at T_c since

$$\frac{\langle M^4 \rangle_L}{3 \langle M^2 \rangle_L^2} = \frac{L^{-\frac{4\beta}{\nu}} F_{M4} \left[(T - T_c) L^{\frac{1}{\nu}} \right]}{\left\{ L^{-\frac{2\beta}{\nu}} F_{M2} \left[(T - T_c) L^{\frac{1}{\nu}} \right] \right\}^2} = F_C \left[(T - T_c) L^{\frac{1}{\nu}} \right]. \quad (1.85)$$

At the critical temperature ($T = T_c$), the scaling function F_C , which is nothing but the ratio of two other scaling functions, is a system size independent constant. As shown in the left panel of Fig. 1.21, for $T > T_c$, the magnetization is described by a Gaussian distribution

$$P_L(M) = \sqrt{\frac{L^d}{\pi \sigma_L}} \exp \left[-\frac{M^2 L^d}{\sigma_L} \right], \quad (1.88)$$

with $\sigma_L = k_B T 2\chi_L$. Since the fourth moment equals three times the second moment squared, i.e.,

$$\langle M^4 \rangle_L = 3 \langle M^2 \rangle_L^2, \quad (1.89)$$

it follows that $U_L(T)$ must be zero for $T > T_c$.

Below the critical temperature ($T < T_c$), there exist one ground state with positive and one with negative magnetization and the corresponding distribution is given by

$$P_L(M) = \frac{1}{2} \sqrt{\frac{L^d}{\pi\sigma_L}} \left\{ \exp \left[-\frac{(M - M_S)^2 L^d}{\sigma_L} \right] + \exp \left[-\frac{(M + M_S)^2 L^d}{\sigma_L} \right] \right\}, \quad (1.91)$$

as illustrated in the right panel of Fig. 1.21.

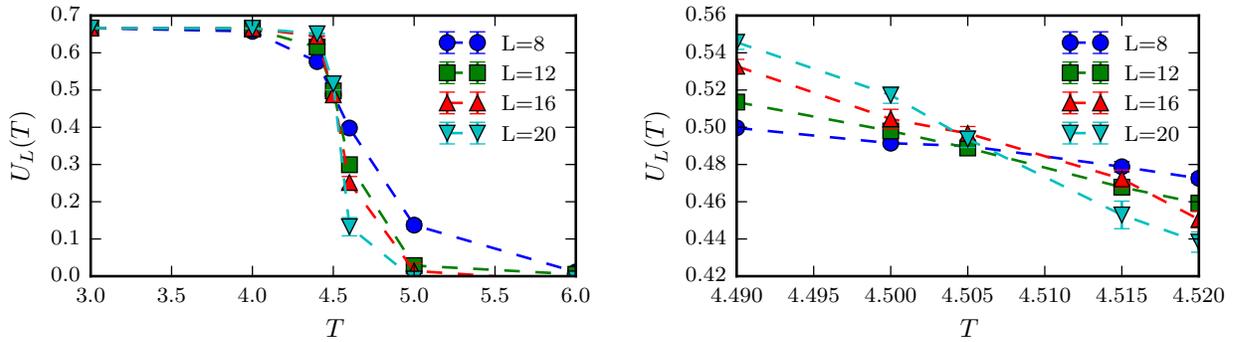


Figure 1.20: The temperature dependence of the Binder cumulant in the three dimensional Ising model for different sizes.

For this distribution, it holds that $\langle M^4 \rangle_L = \langle M^2 \rangle_L^2$ and therefore $U_L(T) = \frac{2}{3}$ for $T < T_c$. In summary, we demonstrated that

$$U_L(T) = \begin{cases} \frac{2}{3} & \text{for } T < T_c \\ \text{const.} & \text{for } T = T_c \\ 0 & \text{for } T > T_c \end{cases} \quad (1.92)$$

The behavior of the Binder cumulant according to Eq. (1.92) is shown in Fig. 1.20 for the three dimensional Ising model and different system sizes. With Eq. (1.92), we have found a very efficient way to calculate the critical temperature based on the temperature sensitive Binder cumulant U_L . For infinite systems, the cumulant exhibits a jump at T_c . All our discussions related to the behavior of different thermodynamic quantities at the critical temperature assumed a pure power law behavior as, for example, the one in Eqs. (1.19), (1.28) and (1.29). However, we would observe such a power law dependence only for temperature values very close to the critical temperature. Far away

from T_c we cannot observe a clear power law behavior anymore, and corrections to scaling have to be employed, i.e.,

$$M(T) = A_0 (T_c - T)^\beta + A_1 (T_c - T)^{\beta_1} + \dots, \quad (1.93)$$

$$\xi(T) = C_0 (T_c - T)^{-\nu} + C_1 (T_c - T)^{-\nu_1} + \dots, \quad (1.94)$$

with $\beta_1 > \beta$ and $\nu_1 < \nu$. These corrections are very important for high quality data, where the errors are small and the deviations become visible. The scaling functions must also be generalized as

$$M(T, L) = L^{-\frac{\beta}{\nu}} F_M \left[(T - T_c) L^{\frac{1}{\nu}} \right] + L^{-x} F_M^1 \left[(T - T_c) L^{\frac{1}{\nu}} \right] + \dots \quad (1.95)$$

with $x = \max \left[\frac{\beta_1}{\nu}, \frac{\beta}{\nu_1}, \frac{\beta}{\nu} - 1 \right]$.

1.5.2 First Order Transition

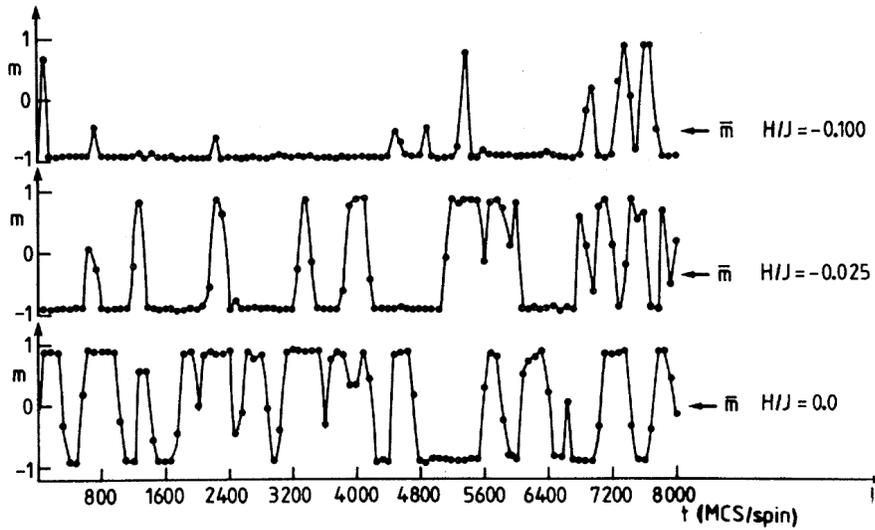


Figure 1.21: The magnetization exhibits a switching behavior if the field vanishes. For non-zero magnetic fields, the magnetization is driven in the direction of the field. The figure is taken from Ref. [23].

Until now, we only focused on the temperature dependence of thermodynamic quantities in the context of the second order phase transition of the Ising model, which is characterized by a discontinuity in the second derivative of the free energy $F = -k_B T \ln(Z)$ (e.g., the susceptibility at the critical temperature). In general, phase transitions are characterized by their *order* which refers to the n^{th} derivative of the free energy not being continuous. In particular, a n^{th} order phase transitions is defined by the lowest non-continuous derivative being of order n . For $T < T_c$, the Ising model exhibits a jump in the magnetization at $H = 0$ which is proportional to the first derivative of the free energy. This leads to a jump in the susceptibility at $H = 0$.

Binder showed that the magnetization as a function of the field H is described by $\tanh(\alpha L^d)$ if the distribution of the magnetization is

given by Eq. (1.91) [23]. Specifically, we find for the magnetization and susceptibility

$$M(H) = \chi_L^D H + M_L \tanh(\beta H M_L L^d), \quad (1.96)$$

$$\chi_L(H) = \frac{\partial M}{\partial H} = \chi_L^D + \frac{\beta M_L L^d}{\cosh^2(\beta H M_L L^d)}. \quad (1.97)$$

Similarly, to the scaling of a second order transition, we can scale the maximum of the susceptibility ($\chi_L(H=0) \propto L^d$) and the width of the peak ($\Delta\chi_L \propto L^{-d}$). To summarize, a first order phase transition is characterized by

1. A bimodal distribution of the order parameter,
2. stochastic switching between the two states in small systems,
3. hysteresis of the order parameter when changing the field,
4. a scaling of the order parameter, or response function according to Eq. (1.97).

1.6 Cluster Algorithms

Based on our discussion in Secs. 1.4.8 and 1.4.9, we have learned that large system sizes lead to longer simulation times—in particular in the vicinity of T_c . All algorithms we considered so far are based on single spin flips only. As a consequence of this update scheme and the critical slowing down, we have to wait sufficiently long between adding a thermodynamic quantity to our averaging procedure to obtain statistically independent samples. The aim of cluster algorithms is to reduce the computation time by flipping multiple spins at the same time. It is essential that the group of spins to flip is chosen with a small acceptance probability. To do this, we will generalize the Ising model and adapt this generalization to our needs.

1.6.1 Potts Model

The Potts model is a generalization of the Ising model to a model with $q \geq 2$ states. It is a very versatile model due to its applications in many fields including sociology, biology and material science.

The Hamiltonian of the system is defined as

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \delta_{\sigma_i \sigma_j} - H \sum_i \sigma_i, \quad (1.98)$$

where $\sigma_i \in \{1, \dots, q\}$ and $\delta_{\sigma_i \sigma_j}$ is unity when nodes i and j are in the same state. The Potts model exhibits a *first order transition* at the critical temperature in two dimensions for $q > 4$, and for $q > 2$ for dimensions larger than two. For $q = 2$, the Potts model is equivalent

to the Ising model. Moreover, there exists a connection between the Potts model and the well known bond percolation model. Kasteleyn and Fortuin demonstrated that the two models have related partition functions [24].

Any thermodynamic system is characterized by its partition function from which all thermodynamic quantities can be derived. Therefore, the partition function completely describes a given thermodynamic system. If two systems have exactly the same partition function, those two systems are equivalent. Using this knowledge, we will prove a relation between the Potts model and bond percolation in the following section.

1.6.2 The Kasteleyn and Fortuin Theorem

We consider the Potts model not on a square lattice but on an arbitrary graph of nodes connected with bonds v . Each node has q possible states and each connection leads to an energy cost of unity if two connected nodes are in a different state and of zero if they are in the same state, i.e.,

$$E = J \sum_v \epsilon_v \quad \text{with} \quad \epsilon_v = \begin{cases} 0 & \text{if endpoints are in the same state} \\ 1 & \text{otherwise} \end{cases} \quad (1.99)$$

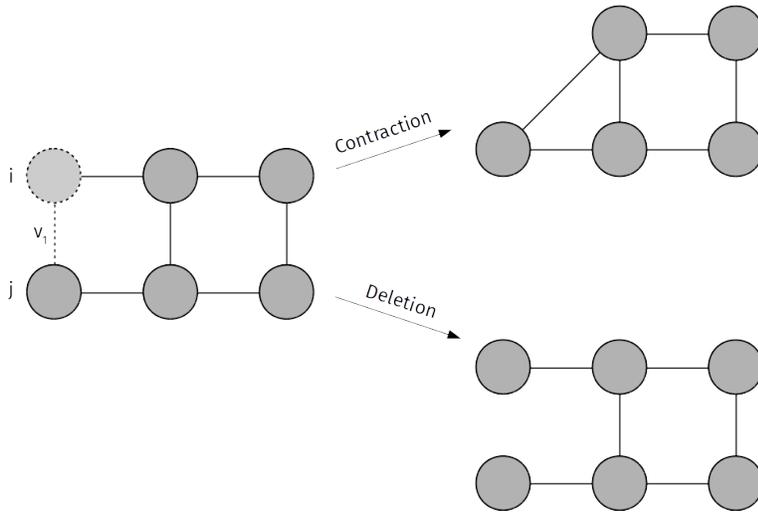


Figure 1.22: Contraction and deletion on a graph.

As depicted in Fig. 1.22, we also account for *contraction* and *deletion* operations on the graph. If two sites are in the same state, they can be merged (contraction) or simply be removed (deletion).

The partition function is the sum over all possible configurations

weighted by the Boltzmann factor and thus given by

$$Z = \sum_X e^{-\beta E(X)} \stackrel{(1.99)}{=} \sum_X e^{-\beta J \sum_v \epsilon_v} = \sum_X \prod_v e^{-\beta J \epsilon_v}. \quad (1.100)$$

We now consider a graph where bond ν_1 connects two nodes i and j with states σ_i and σ_j , respectively. If we would delete bond ν_1 , the partition function is

$$Z_D = \sum_X \prod_{v \neq \nu_1} e^{-\beta J \epsilon_v}. \quad (1.101)$$

We can thus rewrite Eq. (1.100) as

$$\begin{aligned} Z &= \sum_X e^{-\beta J \epsilon_{\nu_1}} \prod_{v \neq \nu_1} e^{-\beta J \epsilon_v} \\ &= \sum_{X: \sigma_i = \sigma_j} \prod_{v \neq \nu_1} e^{-\beta J \epsilon_v} + e^{-\beta J} \sum_{X: \sigma_i \neq \sigma_j} \prod_{v \neq \nu_1} e^{-\beta J \epsilon_v}, \end{aligned}$$

where the first part is the partition function of the contracted graph Z_C and the second part is given by the identity

$$\sum_{X: \sigma_i \neq \sigma_j} \prod_{v \neq \nu_1} e^{-\beta J \epsilon_v} = \sum_X \prod_{v \neq \nu_1} e^{-\beta J \epsilon_v} - \sum_{X: \sigma_i = \sigma_j} \prod_{v \neq \nu_1} e^{-\beta J \epsilon_v} = Z_D - Z_C. \quad (1.102)$$

Summarizing the last results, we find

$$Z = Z_C + e^{-\beta J} (Z_D - Z_C) = p Z_C + (1 - p) Z_D, \quad (1.103)$$

where $p = 1 - e^{-\beta J}$. To be more precise, we expressed the partition function Z as the contracted and deleted partition functions at bond ν_1 . We apply the last procedure to another bond ν_2 and find

$$Z = p^2 Z_{C_{\nu_1, C_{\nu_2}}} + p(1 - p) Z_{C_{\nu_1, D_{\nu_2}}} + (1 - p)p Z_{D_{\nu_1, C_{\nu_2}}} + (1 - p)^2 Z_{D_{\nu_1, D_{\nu_2}}}. \quad (1.104)$$

After applying these operations to every bond, the graph is reduced to a set of separated points corresponding to clusters of nodes which are connected and in the same state out of q states. The partition function reduces to

$$Z = \sum_{\substack{\text{configurations of} \\ \text{bond percolation}}} q^{\# \text{ of clusters}} p^c (1 - p)^d = \left\langle q^{\# \text{ of clusters}} \right\rangle_b, \quad (1.105)$$

where c and d are the numbers of contracted and deleted bonds respectively. In the limit of $q \rightarrow 1$, one obtains the partition function of bond percolation⁹. We found a fundamental relation between a purely geometrical model (bond percolation) and a magnetic model

⁹ In bond percolation, an edge of a graph is occupied with probability p and vacant with probability $1 - p$.

described by a Hamiltonian (Potts model) [25]. Interestingly, we can now choose non-integer values for q . This was meaningless in the original definition of the Potts model since q was introduced as the state of each site.

1.6.3 Coniglio-Klein Clusters

We should bear in mind that the equivalence between bond percolation and the Potts model is of statistical nature since a particular bond configuration may correspond to several spin configurations and vice versa. The fact that the actual spin values are absent in Eq. (1.105) forms the basis for cluster algorithms.

The probability of a given cluster C to be in a certain state σ_0 is independent of the state itself, i.e.,

$$p(C, \sigma_0) = p^{c_C} (1-p)^{d_C} \sum_{\substack{\text{bond percolation} \\ \text{without cluster } C}} q^{\# \text{ of clusters}} p^c (1-p)^d. \quad (1.106)$$

This implies that flipping this particular cluster has no effect on the partition function (and therefore the energy) so that it is possible to accept the flip with probability one. This can be seen by looking at the detailed balance condition of the system

$$p(C, \sigma_1) W[(C, \sigma_1) \rightarrow (C, \sigma_2)] = p(C, \sigma_2) W[(C, \sigma_2) \rightarrow (C, \sigma_1)] \quad (1.107)$$

and using $p(C, \sigma_1) = p(C, \sigma_2)$.

We then obtain for the acceptance probabilities

$$\begin{aligned} A[(C, \sigma_2) \rightarrow (C, \sigma_1)] &= \min \left[1, \frac{p(C, \sigma_2)}{p(C, \sigma_1)} \right] = 1 \quad \text{M(RT)}^2, \\ A[(C, \sigma_2) \rightarrow (C, \sigma_1)] &= \frac{p(C, \sigma_2)}{p(C, \sigma_1) + p(C, \sigma_2)} = \frac{1}{2} \quad \text{Glauber.} \end{aligned} \quad (1.108)$$

Based on these insights, we introduce cluster algorithms which are much faster than single-spin flip algorithms and less prone to the problem of critical slowing down.

1.6.4 Swendsen-Wang Algorithm

The Swendsen-Wang algorithm is a refined Monte Carlo technique which uses the advantage of updating whole clusters of spins. For a certain configuration, we iterate over all bonds connecting spins. Whenever two bonded sites are in the same state, the two sites belong to the same cluster with probability $p = 1 - e^{-\beta J}$. Once the clusters are determined, they can be flipped using any of the updating schemes mentioned before. The basic procedure is as follows:

Swendsen-Wang algorithm

- Occupy the bonds with probability $p = 1 - e^{-\beta J}$ if sites are in the same state.
- Identify the clusters with the Hoshen-Kopelman algorithm.
- Flip the clusters with probability 1/2 for Ising or always choose a new state for $q > 2$.
- Repeat the procedure.

1.6.5 Wolff Algorithm

The Wolff algorithm uses a recursive method to identify clusters and is based on the following steps:

Wolff algorithm

- Choose a site randomly.
- If the neighboring sites are in the same state, add them to the cluster with probability $p = 1 - e^{-\beta J}$.
- Repeat this for any site on the boundaries of the cluster, until all the bonds of the cluster have been checked exactly once.
- Choose a new state for the cluster.
- Repeat the procedure.

1.6.6 Other Ising-like Models

Before focusing on further simulation methods, we briefly discuss other generalizations of the Ising model. The Potts model is only one example of a related physical system. In fact, there exists a large number of models which are modified versions of the Ising model to describe related physical phenomena such as antiferromagnetism, spin glasses or metamagnetism. One of the possible generalizations of the Ising model is the so called n -vector model. Unlike the Potts model, it describes spins as vectors with n components. This model has applications in modeling magnetism or the Higgs mechanism.

The Hamiltonian resembles the one of the Potts model in the sense that it favors spin alignment

$$\mathcal{H} = -J \sum_{\langle i,j \rangle} \vec{S}_i \cdot \vec{S}_j + \vec{H} \sum_i \vec{S}_i. \quad (1.109)$$

with $\vec{S}_i = (S_i^1, S_i^2, \dots, S_i^n)$ and $|\vec{S}_i| = 1$. For $n = 1$ we obtain the Ising model, the case where $n = 2$ corresponds to the *XY-model*, for $n = 3$ we find the *Heisenberg model* and finally, for $n = \infty$ resembles the *spherical model*. Note that the models for various values of n are not equivalent.

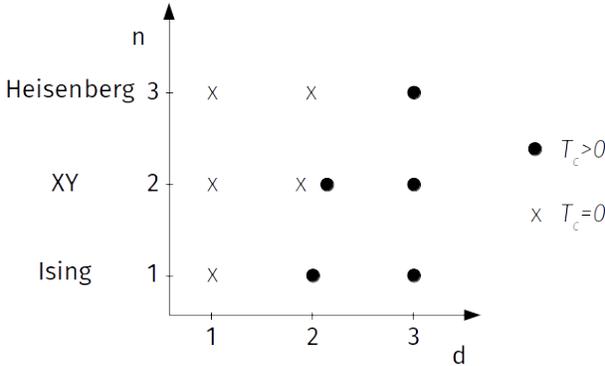


Figure 1.23: The dependence of the critical temperature on the number of vector components n .

In fact, there are huge differences. As an example, the XY-model does not exhibit a phase transition for a non-zero critical temperature from an ordered to a disordered state in two dimensions. The proof of this statement can be found in Ref. [26], and it is known as the Mermin-Wagner theorem. Ernst Ising himself proved in his doctoral thesis that the one dimensional Ising model also does not exhibit a phase transition for $T_c > 0$. In three dimensions, however, both the XY and the Heisenberg model exhibit a phase transition for a non-zero T_c . We summarized these results in Fig. 1.23. The scaling behavior of these models close the critical temperature is very similar to the behavior of the Ising model.

For Monte Carlo simulations with vector-valued spins we have to adapt our simulation methods. The classical strategy is to flip spins by modifying the spin locally trough adding a small $\Delta\vec{S}$ such that $\vec{S}'_i = \vec{S}_i + \Delta\vec{S}$ and $\Delta\vec{S} \perp \vec{S}_i$. The classical Metropolis algorithm can then be used in the same fashion as in the Ising model.

In order to use cluster methods one can project a group of spins onto a plane, and then reflect the spins with respect to the plane. In order to use cluster algorithms one has to find a method to identify equal spins. The probability to find equal spins in a vector-valued model becomes very small one can therefore consider a certain range of values instead.

1.7 Histogram Methods

For computing the thermal average defined by Eq. (1.14), we need to sample different configurations at different temperatures. Another possibility would be to determine an average at a certain tempera-

ture T_0 and extrapolate to another temperature T . In the case of a canonical ensemble, an extrapolation can be achieved by reweighing the histogram of energies $p_{T_0}(E)$ with the Boltzmann factor $\exp(E/T - E/T_0)$. Such histogram methods have first been described in Ref. [27]. We now reformulate the computation of the thermal average of a quantity Q and of the partition function as a sum over all possible energies instead of over all possible configurations and find

$$Q(T_0) = \frac{1}{Z_{T_0}} \sum_E Q(E) p_{T_0}(E) \quad \text{with} \quad Z_{T_0} = \sum_E p_{T_0}(E), \quad (1.110)$$

where $p_{T_0}(E) = g(E) e^{-\frac{E}{k_B T_0}}$ with $g(E)$ defining the *degeneracy of states*, i.e., the number of states with energy E . This takes into account the fact that multiple configurations can have the same energy. The goal is to compute the quantity Q at another temperature T

$$Q(T) = \frac{1}{Z_T} \sum_E Q(E) p_T(E). \quad (1.111)$$

The degeneracy of states contains all the information needed. Using the definition of $g(E)$ yields

$$p_T(E) = g(E) e^{-\frac{E}{k_B T}} = p_{T_0}(E) \exp\left[-\frac{E}{k_B T} + \frac{E}{k_B T_0}\right] \quad (1.112)$$

and with $f_{T_0,T}(E) = \exp\left[-\frac{E}{k_B T} + \frac{E}{k_B T_0}\right]$ we finally obtain

$$Q(T) = \frac{\sum_E Q(E) p_{T_0}(E) f_{T_0,T}(E)}{\sum_E p_{T_0}(E) f_{T_0,T}(E)}. \quad (1.113)$$

With Eq. (1.113) we found a way to compute the value of a quantity Q at any temperature T based on a sampling at a certain temperature T_0 . The drawback of this method is that the values of $Q(E)$ are sampled around the maximum of $p_T(E)$, which converges to a delta distribution for large systems as shown in Fig. 1.24. This means that the statistics are very poor for values of T_0 and T which are substantially different, and results might be inaccurate or even wrong. One possible solution is to interpolate data from several temperatures (multicanonical method) but this involves computations for many temperatures which is also inefficient for large systems. Another solution to this problem, the so-called *broad histogram method*, has been presented in Ref. [28].

1.7.1 Broad Histogram Method

The aim of the broad histogram method is to directly calculate the degeneracy of states over a broader energy range. This concept is

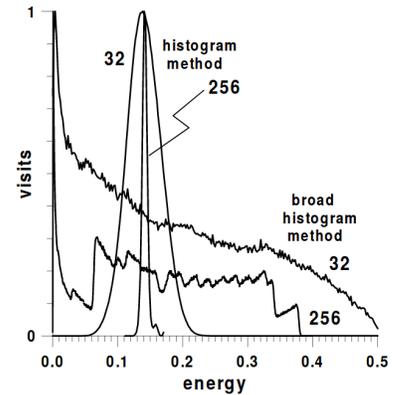


Figure 1.24: An example of the histogram and the broad histogram method for different system sizes. The figure is taken from Ref. [28].

shown in Fig. 1.24. We therefore need to define a Markov process in energy space which is not only exploring regions of certain energies.

Let N_{up} and N_{down} be the numbers of processes which lead to an *increasing* and *decreasing* energy, respectively. Furthermore, we have to keep in mind that the degeneracy of states increases exponentially with energy E , because the number of possible configurations increases with energy. To explore all energy regions equally, we find a condition equivalent to the one of detailed balance, i.e.,

$$g(E + \Delta E) N_{\text{down}}(E + \Delta E) = g(E) N_{\text{up}}(E). \quad (1.114)$$

The motion in phase space towards higher energies can then be penalized with a Metropolis-like dynamics:

- Choose a new configuration,
- if the new energy is lower, accept the move,
- if the new energy is higher then accept with probability $\frac{N_{\text{down}}(E + \Delta E)}{N_{\text{up}}(E)}$.

We obtain the function $g(E)$ by taking the logarithm of Eq. (1.114) and divide by ΔE

$$\log [g(E + \Delta E)] - \log [g(E)] = \log [N_{\text{up}}(E)] - \log [N_{\text{down}}(E + \Delta E)]. \quad (1.115)$$

In the limit of small energy differences, we can approximate the last equation by

$$\frac{\partial \log [g(E)]}{\partial E} = \frac{1}{\Delta E} \log \left[\frac{N_{\text{up}}(E)}{N_{\text{down}}(E + \Delta E)} \right] \quad (1.116)$$

which we can numerically integrate to obtain $g(E)$. Distributions of N_{up} and N_{down} can be obtained by keeping track of these numbers for each configuration at a certain energy. In addition, we also need to store the values of the quantity $Q(E)$ we wish to compute as a thermal average according to

$$Q(T) = \frac{\sum_E Q(E) g(E) e^{-\frac{E}{k_B T}}}{\sum_E g(E) e^{-\frac{E}{k_B T}}}. \quad (1.117)$$

Based on a known degeneracy of states $g(E)$, we can now compute quantities at any temperature.

1.7.2 Flat Histogram Method

Another simulation method which aims at obtaining a broader energy sampling has been described in Ref. [29] and is based on the following procedure:

Flat histogram method

- Start with $g(E) = 1$ and set $f = e$.
- Make a Monte Carlo update with $p(E) = 1/g(E)$.
- If the attempt is successful at E : $g(E) \leftarrow f \cdot g(E)$.
- Obtain a histogram of energies $H(E)$.
- If $H(E)$ is flat enough, then $f \leftarrow \sqrt{f}$.
- Stop when $f \leq 1 + 10^{-8}$.

By setting the Monte Carlo update probability to $1/g(E)$, we obtain smaller transition probabilities for larger energies due to the larger number of possible configurations at the same energy. Therefore the method tends towards energies with fewer configurations. After a successful update, the values of g and of the energy E have to be updated. Once a histogram has been obtained one can increase the precision by decreasing f . The flatness of the histogram can be measured as the ratio of the minimum to the maximum value. For further details, see the *Wang-Landau* algorithm [30, 31].

1.7.3 Umbrella Sampling

The Umbrella sampling technique was developed and proposed in Ref. [32]. The aim is to overcome the problem of the missing ergodicity for certain energy landscapes. As an example, in the Ising model the system could have difficulties in jumping from a positive to a negative magnetization or vice versa if the system is very large. The basic idea is to multiply transition probabilities with a function that is large at the free energy barrier and to later on remove this correction in the averaging step.

$$\tilde{p}(C) = \frac{w(C) e^{-\frac{E(C)}{k_B T}}}{\sum_C w(C) e^{-\frac{E(C)}{k_B T}}} \quad \text{with} \quad \langle A \rangle = \frac{\langle A/w \rangle_w}{\langle 1/w \rangle_w}. \quad (1.118)$$

Summarizing, some of the most common techniques related to the histogram methods are

- Wang-Landau method [30, 31],
- Multiple histogram method [33],
- Multicanonical Monte Carlo [34],
- Flat Histogram method [29],
- Umbrella sampling [32].

1.8 Renormalization Group

In this section, we discuss renormalization group methods and the importance of symmetries to improve our simulations. In particular, we introduce the main concepts of renormalization theory and present some numerical results. For more details we refer to Ref. [35]. Usually, the more information there is available about a system, the better certain quantities can be computed. Close to critical points, changes in the scale of the system can be used to better extrapolate the values to an infinite system. Furthermore, it is possible to develop scale-invariant theories to describe the properties of Ising or similar systems in the vicinity of their critical points.

In the case of self-similar patterns, the invariance under scale transformations is obvious. A curve described by the function $f(x)$ is said to be scale-invariant if $f(\lambda x) = \lambda^\Delta f(x)$. In the case of the Koch curve shown in Fig. 1.25, we obtain the same curve ($\Delta = 1$) after a rescaling with $\lambda = 1/3^n$ with n being an integer. Here, we also want to generalize this intuitive treatment of scale changes to concepts from statistical physics. One possibility is to look at the free energy density and its invariance. To renormalize a system means to change its scale by a factor l such that $\tilde{L} = L/l$. This can be done either in position, or in momentum space.

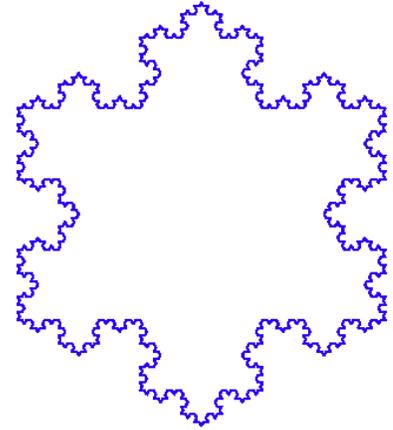


Figure 1.25: The Koch snowflake as an example of a self-similar pattern.

1.8.1 Real Space Renormalization

If a system is invariant under a certain transformation, we can apply this transformation infinitely often without changing the observables of the system. At the critical point, there exist no finite correlation length scales and the properties of the system are invariant under scale changes. We thus regard criticality as a *fixed point* under such renormalization transformations [36]. In order to put the concept of renormalization into a mathematical framework, we discuss the examples of free energy renormalization and of decimation of the one-dimensional Ising model¹⁰. We then generalize the concept in Sec. 1.8.5 and present the implementation of renormalization within Monte Carlo procedures in Sec. 1.8.6.

1.8.2 Renormalization and Free Energy

To build some intuition for renormalization approaches, we consider a scale transformation of the characteristic length L of our system that leads to a rescaled characteristic length $\tilde{L} = L/l$. Moreover, we consider the partition function of an Ising system as defined by the Hamiltonian given in Eq. (1.16). A scale transformation with $\tilde{L} = L/l$

¹⁰ H. J. Maris and L. P. Kadanoff describe in *Teaching the renormalization group* that “communicating exciting new developments in physics to undergraduate students is of great importance. There is a natural tendency for students to believe that they are a long way from the frontiers of science where discoveries are still being made.” [37]

leaves the partition function

$$Z = \sum_{\{\sigma\}} e^{-\beta \mathcal{H}} \quad (1.119)$$

and the corresponding free energy invariant [4]. We therefore find for the free energy densities

$$f(\epsilon, H) = l^{-d} \tilde{f}(\tilde{\epsilon}, \tilde{H}). \quad (1.120)$$

We set $\tilde{\epsilon} = l^{y_T} \epsilon$ and $\tilde{H} = l^{y_H} H$ to obtain

$$\tilde{f}(\tilde{\epsilon}, \tilde{H}) = \tilde{f}(l^{y_T} \epsilon, l^{y_H} H). \quad (1.121)$$

Since renormalization also affects the correlation length

$$\tilde{\zeta} \propto |T - T_c|^{-\nu} = |\epsilon|^{-\nu} \quad (1.122)$$

we can relate the critical exponent ν to y_T . The renormalized correlation length $\tilde{\zeta} = \zeta/l$ scales as

$$\tilde{\zeta} \propto \tilde{\epsilon}^{-\nu}. \quad (1.123)$$

And due to

$$l^{y_T} \epsilon = \tilde{\epsilon} \propto \tilde{\zeta}^{-\frac{1}{\nu}} = \left(\frac{\zeta}{l}\right)^{-\frac{1}{\nu}} \propto \epsilon l^{\frac{1}{\nu}}, \quad (1.124)$$

we find $y_T = 1/\nu$.

The critical point is a fixed point of the transformation since $\epsilon = 0$ at T_c and ϵ does not change independent of the value of the scaling factor.

1.8.3 Majority Rule

A straightforward example which can be regarded as renormalization of spin systems is the *majority rule*. Instead of considering all spins in a certain neighborhood separately, one just takes the direction of the net magnetization of these regions as new spin value, i.e.,

$$\tilde{\sigma}_i = \text{sign} \left(\sum_{\text{region}} \sigma_i \right). \quad (1.125)$$

One has to be careful by applying this transformation. For example, in a one-dimensional lattice with spins pointing up and down it would be an error to apply this transformation on an even number of spins. This may lead to renormalized spin values with value zero. The fact that one deals with system of finite size is also something that has to be taken into account. We can only renormalize up to a certain scale, before finite size effects are visible.

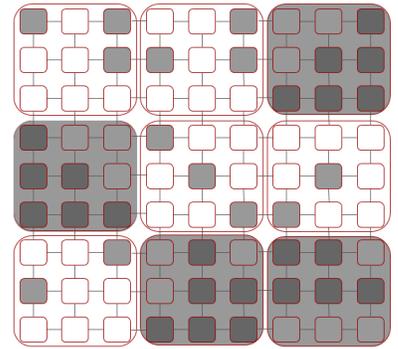


Figure 1.26: An illustration of the majority rule renormalization.

1.8.4 Decimation of the One-dimensional Ising Model

Another possible rule is *decimation* which eliminates certain spins, generally in a regular pattern. As a practical example, we consider the one-dimensional Ising model.

The spins only interact with their nearest neighbors and the coupling constant $K = J/(k_B T)$ is the same for all spins. An example



Figure 1.27: An example of a one-dimensional Ising chain.

of such a spin chain is shown in Fig. 1.27. To further analyze this system, we compute its partition function Z and obtain

$$\begin{aligned}
 Z &= \sum_{\{\sigma\}} e^{K \sum_i \sigma_i} = \sum_{\sigma_{2i}=\pm 1} \prod_{2i} \left[\sum_{\sigma_{2i+1}=\pm 1} e^{K(\sigma_{2i}\sigma_{2i+1} + \sigma_{2i+1}\sigma_{2i+2})} \right] \\
 &= \sum_{\sigma_{2i}=\pm 1} \prod_{2i} \{2 \cosh [K (\sigma_{2i} + \sigma_{2i+2})]\} \\
 &= \sum_{\sigma_{2i}=\pm 1} \prod_{2i} z(K) e^{\tilde{K} \sigma_{2i} \sigma_{2i+2}} \\
 &= [z(K)]^{\frac{N}{2}} \sum_{\sigma_{2i}=\pm 1} \prod_{2i} e^{\tilde{K} \sigma_{2i} \sigma_{2i+2}},
 \end{aligned} \tag{1.126}$$

where we used in the third step that the $\cosh(\cdot)$ function only depends on even spins.

According to Eq. (1.126), the relation

$$Z(K, N) = [z(K)]^{\frac{N}{2}} Z(\tilde{K}, N/2) \tag{1.127}$$

holds as a consequence of the decimation method. The function $z(K)$ is the spin-independent part of the partition function and \tilde{K} is the renormalized coupling constant.

We compute the relation $z(K) e^{\tilde{K} s_{2i} s_{2i+2}} = 2 \cosh [K (s_{2i} + s_{2i+2})]$ explicitly and find

$$z(K) e^{\tilde{K} s_{2i} s_{2i+2}} = \begin{cases} 2 \cosh (2K) & \text{if } s_{2i} = s_{2i+2}, \\ 2 & \text{otherwise.} \end{cases} \tag{1.128}$$

Dividing and multiplying the last two expressions yields

$$e^{2\tilde{K}} = \cosh (2K) \quad \text{and} \quad z^2(K) = 4 \cosh (2K). \tag{1.129}$$

And the renormalized coupling constant \tilde{K} in terms of K is given by

$$\tilde{K} = \frac{1}{2} \ln [\cosh (2K)]. \tag{1.130}$$

We illustrate the fixed point iteration of Eq. (1.130) in Fig. 1.28. We

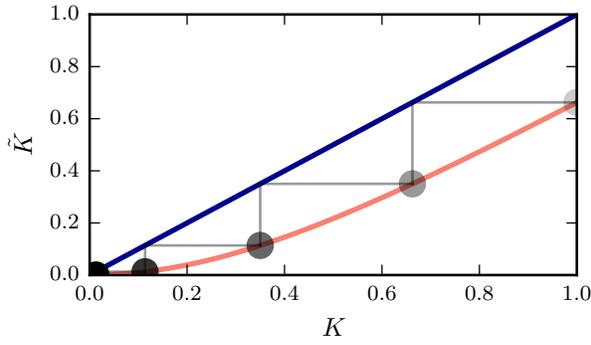


Figure 1.28: An illustration of the fixed point iteration defined by Eq. (1.130).

have thus obtained a rule which describes the change of the coupling constant under renormalization. Given the partition function, we now compute the free energy according to $F = -k_B T N f(K) = -k_B T \ln(Z)$ with $f(K)$ being the free energy density. Taking the logarithm of Eq. (1.127) yields

$$\ln[Z(K, N)] = Nf(K) = \frac{1}{2}N \ln[z(K)] + \frac{1}{2}Nf(\tilde{K}). \quad (1.131)$$

Based on the last equation, we can derive the following recursive relation for the free energy density

$$f(\tilde{K}) = 2f(K) - \ln[2 \cosh(2K)^{1/2}]. \quad (1.132)$$

There exists one *stable fixed point* at $K^* = 0$ and another *unstable one* at $K^* \rightarrow \infty$. Every fixed point ($K^* = \tilde{K}$) implies that Eq. (1.132) can be rewritten due to $f(\tilde{K}) = f(K^*)$.

The case of $K^* = 0$ corresponds to the high-temperature limit where the free energy approaches the value

$$F = -Nk_B T f(K^*) = -Nk_B T \ln(2). \quad (1.133)$$

In this case, the entropy dominates the free energy. For $K^* \rightarrow \infty$, the system approaches the low temperature limit and the free energy is given by

$$F = -Nk_B T f(K^*) = -Nk_B T K = -NJ, \quad (1.134)$$

i.e., given by the internal energy.

1.8.5 Generalization

In the case of the decimation process for the one-dimensional Ising system, we found that one renormalized coupling constant was sufficient to perform a renormalization iteration. In general, multiple

coupling constants are necessary as for example in the case of the two-dimensional Ising model. Thus, we have to construct a renormalized Hamiltonian based on multiple renormalized coupling constants, i.e.,

$$\tilde{H} = \sum_{\alpha=1}^M \tilde{K}_{\alpha} \tilde{O}_{\alpha} \quad \text{with} \quad \tilde{O}_{\alpha} = \sum_i \prod_{k \in c_{\alpha}} \tilde{\sigma}_{i+k} \quad (1.135)$$

where c_{α} is the configuration subset over which we renormalize and

$$\tilde{K}_{\alpha}(K_1, \dots, K_M) \quad \text{with} \quad \alpha \in \{1, \dots, M\}. \quad (1.136)$$

It is important to note that using only M interaction terms instead of an infinite number is a truncation, and in fact leads to systematic errors. The accuracy of this method depends on the number of iterations that we want to take into account.

At T_c there exists a fixed point $K_{\alpha}^* = \tilde{K}_{\alpha}(K_1^*, \dots, K_M^*)$. A possible ansatz to solve this problem is the linearization of the transformation. Thus, we compute the Jacobian $T_{\alpha,\beta} = \frac{\partial \tilde{K}_{\alpha}}{\partial K_{\beta}}$ and obtain

$$\tilde{K}_{\alpha} - K_{\alpha}^* = \sum_{\beta} T_{\alpha,\beta}|_{K^*} (K_{\beta} - K_{\beta}^*). \quad (1.137)$$

We can now construct a flow chart of the coupling constant and obtain values for \tilde{K} for each vector $K = (K_1, \dots, K_M)$. To analyze the behavior of the system close to criticality, we consider eigenvalues $\lambda_1, \dots, \lambda_M$ and eigenvectors ϕ_1, \dots, ϕ_M of the linearized transformation defined by Eq. (1.137). The eigenvectors fulfill $\tilde{\phi}_{\alpha} = \lambda_{\alpha} \phi_{\alpha}$ and the fixed point is unstable if $\lambda_{\alpha} > 1$.

The largest eigenvalue dominates the iteration and we can identify the scaling field $\tilde{\epsilon} = l^{\nu} \epsilon$ with the eigenvector of the transformation, and the scaling factor with eigenvalue $\lambda_T = l^{\nu}$. Then, we compute the exponent ν according to

$$\nu = \frac{1}{y_T} = \frac{\ln(l)}{\ln(\lambda_T)}. \quad (1.138)$$

Based on the last equation, we can now calculate critical exponents using the scaling behavior of the system, if the scaling factor l is known.

1.8.6 Monte Carlo Renormalization Group

The implementation of real space renormalization with Monte Carlo techniques was first proposed in Ref. [38] and then reformulated in

Ref. [39].

Since we are dealing with generalized Hamiltonians with many interaction terms, we compute the thermal average of O_α according to

$$\langle O_\alpha \rangle = \frac{\sum_{\{\sigma\}} O_\alpha e^{\sum_\beta K_\beta O_\beta}}{\sum_{\{\sigma\}} e^{\sum_\beta K_\beta O_\beta}} = \frac{\partial F}{\partial K_\alpha} \quad (1.139)$$

where F is the free energy.

Using the fluctuation-dissipation theorem, we can also numerically compute the response functions

$$\begin{aligned} \chi_{\alpha,\beta} &= \frac{\partial \langle O_\alpha \rangle}{\partial K_\beta} = \langle O_\alpha O_\beta \rangle - \langle O_\alpha \rangle \langle O_\beta \rangle, \\ \tilde{\chi}_{\alpha,\beta} &= \frac{\partial \langle \tilde{O}_\alpha \rangle}{\partial K_\beta} = \langle \tilde{O}_\alpha O_\beta \rangle - \langle \tilde{O}_\alpha \rangle \langle O_\beta \rangle. \end{aligned}$$

We also find with Eq. (1.139) that

$$\tilde{\chi}_{\alpha,\beta}^{(n)} = \frac{\partial \langle \tilde{O}_\alpha^{(n)} \rangle}{\partial K_\beta} = \sum_\gamma \frac{\partial \tilde{K}_\gamma}{\partial K_\beta} \frac{\partial \langle \tilde{O}_\alpha^{(n)} \rangle}{\partial K_\gamma} = \sum_\gamma T_{\gamma,\beta} \chi_{\alpha,\gamma}^{(n)}. \quad (1.140)$$

It is thus possible to derive a value of $T_{\gamma,\beta}$ from the correlation functions by solving a set of M coupled linear equations. At point $K = K^*$, we can apply this method in an iterative manner to compute critical exponents as suggested by Eq. (1.138).

There are many error sources in this technique, that originate from the fact that we are using a combination of several tricks to obtain our results:

- Statistical errors,
- Truncation of the Hamiltonian to the M^{th} order,
- Finite number of scaling iterations,
- Finite size effects,
- No precise knowledge of K^* .

1.9 Boltzmann Machine

In recent years the field of machine learning experienced a substantial boom due to the growing amount of training data and the increase in computing power. Many algorithms that had been around for decades have been developed further and new methods emerged during the

past few years. The general idea behind machine learning is the development of computational methods to *perform a certain task* with a performance that can be *improved with more training data*. One could think of the learning problem, in more abstract terms, as a function which depends on multiple parameters and maps a given input to an output. We distinguish between *supervised* and *unsupervised* learning tasks. In supervised learning, we are given input-output pairs and adjust the parameters of the function such that we obtain good performance in mapping given inputs to desired outputs. In the case of unsupervised learning, we aim at extracting the underlying probability distribution of the sample data.

Many learning models and algorithms lack a solid theoretical basis and are therefore not very well understood. Recently, many attempts have been made to characterize certain learning algorithms with methods from statistical physics such as renormalization group approaches [40]. We therefore now focus on one particular example, the so-called *Boltzmann machine*, whose origins lie in statistical physics [41]. This allows us to transfer some of our acquired knowledge on computational sampling techniques in statistical physics to the field of machine learning.

1.9.1 Hopfield Network

We begin our excursion to Boltzmann machines with a network consisting of neurons which are fully connected, i.e., every single neuron is connected to all other neurons. A neuron represents a node of a network and is nothing but a function of I different inputs $\{x_i\}_{i \in \{1, \dots, I\}}$ which are weighted by $\{w_i\}_{i \in \{1, \dots, I\}}$ to compute and output y . A single neuron is shown in Fig. 1.29.

In terms of a Hopfield network, we consider discrete inputs $x_i \in \{-1, 1\}$. The activation of neuron i is given by

$$a_i = \sum_j w_{ij} x_j, \quad (1.141)$$

where we sum over the inputs. The weights fulfill $w_{ij} = w_{ji}$ and $w_{ii} = 0$. Similarly to the Ising model, the associated energy is given by

$$E = -\frac{1}{2} \sum_{i,j} w_{ij} x_i x_j - \sum_i b_i x_i, \quad (1.142)$$

where b_i is the bias term.

The dynamics of a Hopfield network is given by

$$x_i(a_i) = \begin{cases} 1 & \text{if } a_i \geq 0, \\ -1 & \text{otherwise.} \end{cases} \quad (1.143)$$

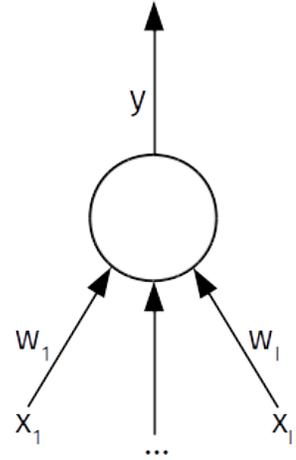


Figure 1.29: An illustration of a single neuron with output y , inputs $\{x_i\}_{i \in \{1, \dots, I\}}$ and weights $\{w_i\}_{i \in \{1, \dots, I\}}$.

The energy difference ΔE_i after neuron i has been updated is

$$\Delta E_i = E(x_i = -1) - E(x_i = 1) = 2 \left(b_i + \sum_j w_{ij} x_j \right). \quad (1.144)$$

We can absorb the bias b_i in the sum by having an extra active unit at every node in the network. We thus showed that the activation defined by Eq. (1.141) equals one half of the energy difference ΔE_i .

Given the update rule defined by Eq. (1.143), the energy of the Hopfield network is decreasing over time towards a value which might not correspond to the global minimum. One application of Hopfield networks is to store certain information in local minima with only setting the values of a small number of neurons. Starting from such an initial configuration, the dynamics converges towards the state containing the desired information.

1.9.2 Boltzmann Machine Learning

For some applications, finding a local minimum based on the deterministic update rule defined by Eq. (1.143) might not be sufficient. Similar to the discussion of Monte Carlo methods for Ising systems, we employ an update probability

$$p_i = \frac{1}{1 + \exp(-\Delta E_i/T)} = \sigma(2a_i/T) \quad (1.145)$$

to set neuron i to unity independent of its state [42]. Here, $\sigma(x) = 1/[1 + \exp(-x)]$ denotes the sigmoid function. As defined in Eq. (1.144), the energy difference ΔE_i is the gap between a configuration with an active neuron i and an inactive one. The parameter T acts as temperature equivalent¹¹.

A closer look at Eqs. (1.142) and (1.145) tells us that we are simulating a Hamiltonian system with Glauber dynamics. Due to the fulfilled detailed balance condition, we reach thermal equilibrium and find again for the probabilities of the system to be in state X or Y ¹²

$$\frac{p_{\text{eq}}(Y)}{p_{\text{eq}}(X)} = \exp\left(-\frac{E(Y) - E(X)}{T}\right). \quad (1.146)$$

Independent of the initial configuration, the last stochastic update procedure always leads to a unique thermal equilibrium configuration fully determined by its energy. We divide the Boltzmann machine units into *visible* and *hidden units* represented by the non-empty set V and the possibly empty set H , respectively. The visible units are set by the environment whereas the hidden units are additional variables which might be necessary to model certain outputs. Let $P'(v)$ be the probability distribution over the visible units v in a freely running

¹¹ For $T \rightarrow 0$, we recover deterministic dynamics as described by Eq. (1.143).

¹² Here we set $k_B = 1$.

network. It can be obtained by marginalizing over the corresponding joint probability distribution, i.e.,

$$P'(v) = \sum_h P'(v, h), \quad (1.147)$$

where h represents a hidden unit. The goal is to come up with a method such that $P'(v)$ approaches the unknown environment distribution $P(v)$. We measure the difference between $P'(v)$ and $P(v)$ in terms of the Kullback-Leibler divergence (relative entropy)

$$G = \sum_{v \in V} P(v) \ln \left[\frac{P(v)}{P'(v)} \right]. \quad (1.148)$$

To minimize G , we perform a gradient descent according to

$$\frac{\partial G}{\partial w_{ij}} = -\frac{1}{T} (p_{ij} - p'_{ij}), \quad (1.149)$$

where p_{ij} is the probability that two units are active on average if the environment is determining the states of the visible units and p'_{ij} is the corresponding probability in a freely running network without a coupling to the environment. Both probabilities are measured at thermal equilibrium. In the literature, the probabilities p_{ij} and p'_{ij} are also often defined in terms of the thermal averages $\langle x_i x_j \rangle_{\text{data}}$ and $\langle x_i x_j \rangle_{\text{model}}$, respectively. The weights w_{ij} of the network are then updated according to

$$\Delta w_{ij} = \epsilon (p_{ij} - p'_{ij}) = \epsilon (\langle x_i x_j \rangle_{\text{data}} - \langle x_i x_j \rangle_{\text{model}}), \quad (1.150)$$

where ϵ is the learning rate. For general learning tasks, the Boltzmann machine is impractical since the times to reach the equilibrium distribution become large for large system sizes. The so-called restricted Boltzmann machine is not taking into account mutual connections within the set of hidden and visible units, and turned out to be more suitable for learning tasks. In the case of restricted Boltzmann machines, the weight update is given by

$$\Delta w_{ij} = \epsilon (\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}}), \quad (1.151)$$

where v_i and h_j represent visible and hidden units, respectively. Instead of sampling the configurations for computing $\langle v_i h_j \rangle_{\text{data}}$ and $\langle v_i h_j \rangle_{\text{model}}$ at thermal equilibrium, we could also just consider a few relaxation steps. This method is called *contrastive divergence* and defined by the following update rule

$$\Delta w_{ij}^{\text{CD}} = \epsilon (\langle v_i h_j \rangle_{\text{data}}^0 - \langle v_i h_j \rangle_{\text{model}}^k), \quad (1.152)$$

where the superscript indices indicate the number of updates.

1.10 Parallelization

After introducing different algorithms for the simulation of different equilibrium systems, we now discuss parallelization techniques to make the necessary computations faster. Parallelization can lead computation times which are entire orders of magnitudes faster compared to a purely serial simulation! The easiest way to parallelize and to obtain better statistics is *farming*, i.e., executing the simulation on different computers or processors with different initial conditions. This technique is useful in situations such as Ising Monte Carlo simulations where one has to repeat the exact same sequence of operations as many times as possible. Farming is, of course, just a method to improve statistics with generating more samples for system sizes that fit into a single machine. More sophisticated parallelization techniques are presented in the subsequent sections¹³.

¹³ The courses High Performance Computing for Science and Engineering I & II provide an in-depth discussion of parallelization methods. A good introduction to parallel programming has also been prepared by the Jülich Supercomputing Centre: download slides.

1.10.1 Multi-Spin Coding

The idea behind *multi-spin coding* is based on the fact that on a 64 bit system not all the bits are used for computations. This method is useful for integer variables limited to a certain value. In the case of the Ising model, we are dealing with spins $\sigma_i \in \{\pm 1\}$. Storing such spin values in 64 bit integers is not only a waste of memory but also leads to a waste of computation time since most of the bits are not carrying any information. In a cubic lattice, the local field takes values $h_i \in \{0, \pm 2, \pm 4, \pm 6\}$ what corresponds to seven different values. In binary representation, three bits are sufficient to store up to eight different numbers. Therefore, three bits would be enough to represent all possible local field and energy values of any site σ_i while 61 bits remain unused. Spin configurations can also be stored more efficiently using the sequences (000) and (001) which correspond to spin down and spin up, respectively. The XOR function (\oplus) of two spin sequences yields (000) whenever two spins are parallel and (001) for anti-parallel spins. We can now store 21 spin values $\sigma_i, \dots, \sigma_{i+20}$ in a single 64 bit integer word, i.e.,

$$N = (\underbrace{\delta_1, \delta_2, \delta_3, \delta_4, \dots, \delta_{62}, \delta_{63}, \delta_{64}}_{\sigma_{i+20}}), \quad (1.153)$$

where $\delta_{(\cdot)} \in \{0, 1\}$. Similarly, we also store the six neighbors of each spin in N in integer words K_1, \dots, K_6 . Thus, only one bit per computer word remains unused, instead of 63. No spins in N shall be neighbors and one has to be very careful with the organization of storing spins. The flipping probabilities for all 21 spins are determined by

$$N \oplus K_1 + N \oplus K_2 + N \oplus K_3 + N \oplus K_4 + N \oplus K_5 + N \oplus K_6. \quad (1.154)$$

After determining the flipping probabilities, each value has to be extracted and compared with a random number. To extract the information of one lattice site, we define the mask $7 = (0, \dots, 0, 1, 1, 1)$. We apply the AND (\wedge) function to Eq. (1.154) and obtain

$$7 \wedge N_j = (0, \dots, 0, 1, 1, 1) \wedge (\delta_1, \dots, \delta_{62}, \delta_{63}, \delta_{64}) = (0, \dots, 0, \tilde{\delta}_{62}, \tilde{\delta}_{63}, \tilde{\delta}_{64}), \quad (1.155)$$

where $\tilde{\delta}_{(\cdot)}$ are the last three bit values of the integer word defined by Eq. (1.154). With the circular right shift operator or the circular left shift operator we can also access the remaining bits of Eq. (1.154). According to Ref. [43] *“multi spin coding makes a program more complicated and error-prone but may save a lot of memory and computer time for large systems. The more complicated the interaction is, the less useful is multispin coding, and for continuous degrees of freedom it does not seem to work at all.”*

1.10.2 Vectorization

Vectorization is technique which allows to perform multiple operations at the same time. It only works in the innermost loops of a program. As an example, we consider the following loop:

```
i_max = 10000;
for (i=1; i<=i_max; i++)
{
    A(i) = B(i) * (C(i) + D(i));
}
```

With the help of vectorizing compilers, the loop can be made more efficient for simulations by executing multiple operations simultaneously. In the case of more complicated routines, this may have to be implemented directly. The drawbacks of vectorization include:

- Multiple short loops,
- conditional branchings like if-statements,
- indirect addressing.

For optimal performance, the instructions must be repetitive without interruptions or exceptions such as if statements. There are ways to handle some cases in which a distinction is needed. An example would be replacing

```
if (P(i)>z)
{
    s = -s;
}
```

by

```
s = s*sign(z-P(i));
```

Moreover, one has to make sure that the inner loop is the largest and loops that cannot be vectorized should be split up. It is also recommended to make use of vectorized random number generators.

1.10.3 Domain Decomposition

Monte Carlo simulations on regular lattices are well-suited for parallelization. The following points summarize the basic ideas behind parallelization:

Domain decomposition

- Nearest-neighbor updates do not involve the whole system.
- Domain decomposition into sublattices is possible.
- The decomposed domains are distributed using MPI.
- Sublattices are extracted with logical masks.
- A periodic shift (CSHIFT) is used to obtain neighbors for periodic boundary conditions.

There are many options for dividing a system into domains. Depending on the situations it can be more convenient to increase the number of domains or reducing the interfaces between domains in case that the processes need to work independently (e.g., in the case of distributed memory). It is also possible to dynamically change the domain sizes and interface positions.

To parallelize a routine one has to use specific programming languages created specifically for parallelizing or embed special libraries in which the parallelization has been implemented in such a way that it can be summoned in standard programming languages such as C++. An interesting example is CUDA (Compute Unified Device Architecture), a parallel computing platform and programming model created by NVIDIA and implemented in their graphics processing units. Moreover, MPI (Message Passing Interface) is a standardized and portable message-passing system. This means that instructions can be passed by the user within programs written in languages such as Java or C++.

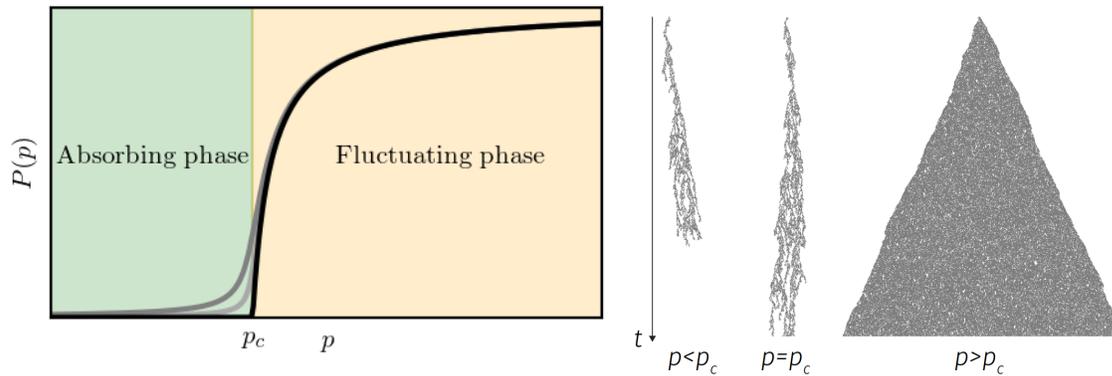
The bottleneck of parallelization is the communication between processors. Processors are not isolated units that completely work on their own, and the more one slices the system into domains, the more communication between processors is needed. This is generally not efficient and usually slows down the parallelization¹⁴.

¹⁴ Furthermore, some parts of a program are not parallelizable. Amdahl's law establishes a relation between the theoretical speed-up and the fraction of parallelizable code.

1.11 Non-Equilibrium Systems

Models approaching thermal equilibrium obey detailed balance, that is their probability flows cancel out each other. With the Ising model and its generalization the Potts model, we studied two examples of equilibrium systems. In Sec. 1.6.2, we have shown that the Potts model also contains isotropic (bond) percolation as a special case. These models have in common that they exhibit phase transitions with universal features, such as scaling laws and universal exponents. In contrast to isotropic percolation, directed percolation defines a non-equilibrium process with different universal properties [4]. Both processes are very fundamental in the sense that they are simple to define, but also include features which are observed in a variety of models and contexts. In particular, models describing certain hadronic interactions or the spreading of opinions or diseases have been mapped to equilibrium (percolation) and non-equilibrium (directed percolation) processes [4, 44].

1.11.1 Directed Percolation



Isotropic percolation is an equilibrium process in which a site or bond is occupied with a certain probability p . However, in the case of directed percolation an additional constraint is defined, namely in each time step occupied sites or bonds occupy their neighbors with probability p only along a given direction [45]. Time is therefore a relevant parameter.

Both models have a critical probability p_c at which a percolating (permeable) phase occurs. This behavior is illustrated in Fig. 1.30. In the vicinity of p_c the percolation order parameter $P(p)$, the fraction of sites of the largest cluster (isotropic percolation) or the density of active sites (directed percolation), takes the form: $P(p) \propto (p - p_c)^\beta$.

Figure 1.30: In the left panel, we show the second-order phase transition of the directed percolation model. The grey lines illustrate the effect of a field-like contribution. An example of bond directed percolation on a two-dimensional square lattice with different control parameter values p is shown in the right panel. At around $p_c = 0.6447$ a second-order phase transition occurs [4].

Although, the free energy is not defined in non-equilibrium processes one can distinguish between first-order and continuous phase transitions based on the behavior of the order parameter.

Another formulation of directed percolation is given by the *contact process*. On a lattice with active ($s_i(t) = 1$) and inactive ($s_i(t) = 0$) sites, we sequentially update the system according to the following dynamics:

Let the number of active neighbors be $n_i(t) = \sum_{\langle i,j \rangle} s_j(t)$ and a new value $s_i(t + dt) \in \{0, 1\}$ is obtained according to the transition rates

$$w[0 \rightarrow 1, n] = (\lambda n)/(2d) \quad \text{and} \quad w[1 \rightarrow 0, n] = 1, \quad (1.156)$$

where d is the dimension of the system.

On a square lattice the critical values λ_c and β for the contact process with $d \leq d_c$ are given in Tab. 1.2. The contact process has the same characteristic critical exponents as directed percolation what implies that they both belong to the same universality class. According to Ref. [4] the directed percolation universality class requires:

1. a continuous phase transition from a fluctuating phase to a unique absorbing state,
2. a transition with a one-component order parameter,
3. local process dynamics,
4. no additional attributes, such as conservation laws or special symmetries.

Table 1.2: The critical infection rate λ_c and the critical exponent β on a square lattice for different dimensions d . At and above the critical dimension $d_c = 4$ the critical exponents are described by mean-field theory [4]. The values are rounded to the second decimal and taken from Refs. [46] and [47].

	$d = 1$	$d = 2$	$d = 3$	$d_c = 4$
λ_c	3.30	1.65	1.32	1.20
β	0.28	0.59	0.78	1

1.11.2 Kinetic Monte Carlo (Gillespie Algorithm)

Non-equilibrium systems are not described by a Hamiltonian and we cannot use the same algorithms which we employed for studying the Ising model or other equilibrium models. Time is now a physical parameter and not a mere virtual property of a Markov chain.

A standard algorithm for simulating non-equilibrium dynamics is the *kinetic Monte Carlo* method which is also often referred to as *Gillespie algorithm*.

To give an example, the kinetic Monte Carlo algorithm is applied to the contact process on a two-dimensional square lattice. Recovery and activation define $n = 2$ processes with corresponding rates $R_1 = 1$ and $R_2 = \lambda$.

At time t , the subpopulation $N_1(t)$ consists of all active nodes which recover with rate R_1 . The total recovery rate is then given by $Q_1(t) = N_1(t)$.

On the square lattice only nearest-neighbor interactions are considered and the total rate of the second process (activation) is obtained by computing n_i and $w [0 \rightarrow 1, n_i]$ according to Eq. (1.156) for all nodes.

The following steps summarize a general kinetic Monte Carlo update¹⁵

Kinetic Monte Carlo algorithm

1. Identify all individual rates (per node) $\{R_i\}_{i \in \{1, \dots, n\}}$.
2. Determine the overall rates (all nodes) $\{Q_i\}_{i \in \{1, \dots, n\}}$, where $Q_i = R_i N_i$ and N_i defines the subpopulation which corresponds to R_i . It is possible that some subpopulations are identical but correspond to different rates.
3. Let $\eta \in [0, 1)$ be a uniformly distributed random number and $Q = \sum_i Q_i$. The process with rate R_j occurs if $\sum_{i=1}^{j-1} Q_i \leq \eta Q < \sum_{i=1}^j Q_i$.
4. Let $\epsilon \in [0, 1)$ be a uniformly distributed random number. Then the time evolves as $t \rightarrow t + \Delta t$, where $\Delta t = -Q^{-1} \log(1 - \epsilon)$ and return to step 2.

¹⁵ It is important to use local updates and not recommended to iterate over the whole system at each time step.

References

- [1] Solé, R. V. *Phase Transitions* (Princeton University Press, 2011).
- [2] Ebeling, W. & Sokolov, I. M. *Statistical thermodynamics and stochastic theory of nonequilibrium systems*, vol. 8 (World Scientific Publishing Company, 2005).
- [3] Gibbs, J. W. On the fundamental formula of statistical mechanics, with applications to astronomy and thermodynamics. In *Proceedings of the American Association for the Advancement of Science*, 57–58 (1884).
- [4] Henkel, M., Hinrichsen, H. & Lübeck, S. *Non-Equilibrium Phase Transitions Volume 1 – Absorbing Phase Transitions* (Springer Science & Business Media, 2008).
- [5] Böttcher, A. & Silbermann, B. *Introduction to large truncated Toeplitz matrices* (Springer Science & Business Media, 2012).
- [6] Yang, C. N. The spontaneous magnetization of a two-dimensional ising model. *Phys. Rev.* **85**, 808 (1952).
- [7] Wolf, W. P. The ising model and real magnetic materials. *Braz. J. Phys.* **30**, 794–810 (2000).

- [8] Onsager, L. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Phys. Rev.* **65**, 117 (1944).
- [9] Ferrenberg, A. M., Xu, J. & Landau, D. P. Pushing the limits of monte carlo simulations for the three-dimensional ising model. *Phys. Rev. E* **97**, 043301 (2018).
- [10] Stanley, H. Scaling, universality, and renormalization: Three pillars of modern critical phenomena (1999). URL http://link.springer.com/chapter/10.1007%2F978-1-4612-1512-7_39.
- [11] Stanley, H. *Introduction to Phase Transitions and Critical Phenomena* (Clarendon, Oxford, 1971).
- [12] Sengers, J., Sengers, J. L. & Croxton, C. Progress in liquid physics. *Wiley, Chichester* (1978).
- [13] Kadanoff, L. Critical behaviour, universality and scaling. In M.S. Green, editor, *Proceedings of the 1970 Varenna summer school on critical phenomena* (1971).
- [14] Pelissetto, A. & Vicari, E. Critical phenomena and renormalization-group theory. *Phys. Rep.* **368**, 549–727 (2002).
- [15] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M., Teller, A. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* (1953). URL http://en.wikipedia.org/wiki/Equation_of_State_Calculations_by_Fast_Computing_Machines.
- [16] Barth, K. *Oral History Transcript — Dr. Marshall Rosenbluth* (2003). URL http://www.aip.org/history/ohilist/28636_1.html.
- [17] Gubernatis, J. Marshall rosenbluth and the metropolis algorithm. *AIP* (2005). URL <http://scitation.aip.org/content/aip/journal/pop/12/5/10.1063/1.1887186>.
- [18] Segré, E. *Enrico Fermi, Physicist* (University of Chicago Press, 1970).
- [19] Hastings, W. Monte carlo sampling methods using markov chains and their applications. *BioMetrika* (1969). URL <http://biomet.oxfordjournals.org/content/57/1/97>.
- [20] Creutz, M. Microcanonical monte carlo simulation. *Phys. Rev. Lett.* **50**, 1411 (1983).
- [21] Vichniac, G. Y. Simulating physics with cellular automata. *Phys. D* **10**, 96–116 (1984).

- [22] Binder, K. & Heermann, D. *Monte Carlo Simulation in Statistical Physics* (Springer, 1997).
- [23] Binder, K. & Landau, D. Finite-size scaling at first-order phase transitions. *Phys. Rev. B* **30**, 1477 (1984).
- [24] Kasteleyn, P. & Fortuin, C. Phase transitions in lattice systems with random local properties. *J. Phys. Soc. Jpn.* **26**, 11 (1969).
- [25] Wu, F. Percolation and the potts model. *J. Stat. Phys.* **18**, 115–123 (1978).
- [26] Mermin, N. & Wagner, H. Absence of ferromagnetism or anti-ferromagnetism in one- or two-dimensional isotropic heisenberg models. *Phys. Rev. Lett.* 1133 (1966).
- [27] Salzburg, Z., Jacobson, J., Fickett, W. & Wood, W. Excitation of non-radial stellar oscillations by gravitational waves: a first model. *Chem. Phys.* 65 (1959).
- [28] de Oliveira, P., Penna, T. & Herrmann, H. Broad histogram method. *Braz. J. Phys.* 677 (1996).
- [29] Wang, F. & Landau, D. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E* **64**, 056101 (2001).
- [30] Wang, F. & Landau, D. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86**, 2050 (2001).
- [31] Zhou, C., Schulthess, T. C., Torbrügge, S. & Landau, D. Wang-landau algorithm for continuous models and joint density of states. *Phys. Rev. Lett.* **96**, 120201 (2006).
- [32] Torrie, G. & Valleau, J. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *J. Comp. Phys.* **23**, 187–199 (1977). URL <http://www.sciencedirect.com/science/article/pii/0021999177901218>.
- [33] Ferrenberg, A. Am ferrenberg and rh swendsen, *phys. rev. lett.* 63, 1195 (1989). *Phys. Rev. Lett.* **63**, 1195 (1989).
- [34] Berg, B. A. Introduction to multicanonical monte carlo simulations. *Fields Inst. Commun.* **26**, 1–24 (2000).
- [35] Niemeijer, T. & van Leeuwen, J. Renormalization: Theory ising-like spin systems. *Phase transitions and critical phenomena* **6**. URL http://en.wikipedia.org/wiki/Phase_transitions_and_critical_phenomena.

- [36] Wilson, K. G. The renormalization group: Critical phenomena and the kondo problem. *Rev. Mod. Phys.* **47**, 773–840 (1975). URL <http://link.aps.org/doi/10.1103/RevModPhys.47.773>.
- [37] Maris, H. J. & Kadanoff, L. P. Teaching the renormalization group. *Am. J. Phys.* **46**, 652–657 (1978).
- [38] Ma, S. Monte carlo renormalization group. *Phys. Rev. Lett.* **37**, 461 (1976).
- [39] Swendsen, R. Monte carlo renormalization group. *Phys. Rev. Lett.* **42**, 859 (1979).
- [40] Mehta, P. & Schwab, D. J. An exact mapping between the variational renormalization group and deep learning. *arXiv preprint arXiv:1410.3831* (2014).
- [41] Hinton, G. E. Deterministic boltzmann learning performs steepest descent in weight-space. *Neural Comp.* **1**, 143–150 (1989).
- [42] Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. A learning algorithm for boltzmann machines. *Cogn. Sci.* **9**, 147–169 (1985).
- [43] Baumgärtner, A. *et al.* *Applications of the Monte Carlo method in statistical physics*, vol. 36 (Springer Science & Business Media, 2013).
- [44] Grassberger, P. On the critical behavior of the general epidemic process and dynamical percolation. *Math. Biosci.* (1983).
- [45] Broadbent, S. & Hammersley, J. Percolation processes I. crystals and mazes. *Proceedings of the Cambridge Philosophical Society* (1957).
- [46] Moreira, A. G. & Dickman, R. Critical dynamics of the contact process with quenched disorder. *Phys. Rev. E* (1996).
- [47] Sabag, M. M. S. & de Oliveira, M. J. Conserved contact process in one to five dimensions. *Phys. Rev. E* (2002).